The Linear Elamite Script

Phonography, Sign Inventory, and the Limits of Decipherment

A Mathematical Analysis

Silvano Salvador

Contents

Abstract **11**

1 The Problem and Its Mathematical Structure **13**
1.1 The Inherited Debate and Its Epistemological Deficits 13
1.2 The Core Mathematical Object: A Formal Description of the Corpus 14
1.3 The Four Analytical Axes and Their Relationship 15
1.4 Formalizing the Hypothesis Space 16
1.5 Relationship to the Companion Volumes 17
1.6 Structure of the Volume 17
1.7 The Three-Century Window: Historical and Archaeological Context 17
1.8 The Writing-System Typology: A Formal Taxonomy 18
1.9 Quantitative Palaeography: Methodology and Scope 19
1.10 The Elamite Language: Phonological and Morphological Background 20

2 Material Evidence and Epigraphic Chronology **22**
2.1 Inscription Supports and Their Chronological Significance 22
2.2 Palaeographic Sign-Form Analysis 23
2.3 Bayesian Chronological Model 24
2.4 The Bilingual Monuments: A Formal Alignment Framework 24
2.5 Spatial Distribution and the Findspot Network 25
2.6 Inscriptions on Metal: Silver, Gold, and Copper Objects 26
2.7 Scribal Workshop Evidence and the Question of Standardization 26
2.8 The Konar Sandal Inscriptions and the Question of Eastern Diffusion 27
2.9 Radiocarbon Evidence and Its Integration with the Epigraphic Chronology . 28

3 Information-Theoretic Profiling of the Corpus **29**
3.1 Unigram Entropy and the Sign-Frequency Distribution 29
3.2 Zipf-Law Analysis 30
3.3 Conditional Entropy and the Phonographic Consistency Diagnostic 30
3.4 Positional Entropy Profiles 31
3.5 Model-Order Ceiling Analysis 32
3.6 Rényi Entropy and the Sign-Concentration Profile 33
3.7 Mutual Information Between Inscription Position and Sign Identity 34
3.8 Diachronic Entropy Analysis: Testing for Temporal Trends 34
3.9 Scribal-Hand Analysis and Individual Variation 35

4 Bigram and Positional Analysis **37**
4.1 The Bigram Transition Matrix 37
4.2 Forbidden Bigrams and the Sign-Compatibility Structure 38
4.3 Cross-Entropy Against Attested Syllabic Scripts 38
4.4 Variable-Order Markov Analysis and Context-Tree Weighting 39

4.5 Transfer Entropy Across Inscription Sub-Groups 39
4.6 Skip-Gram Statistics and Non-Adjacent Co-occurrence 40
4.7 The Sign-Boundary Problem: Determiners, Classifiers, and Multi-Sign Groups 41

5 Reference-Corpus Classification 43
5.1 The Typological Reference Panel 43
5.2 The Proto-Elamite Comparison 44
5.3 The Cypro-Minoan Parallel 44
5.4 The Meroitic and Byblian Syllabaries as Near-Parallel Cases 45
5.5 Formal Criteria for Script-Type Classification 45
5.6 Perplexity Comparison Across the Four Principal Hypotheses 46
5.7 Robustness of the Typological Classification to Reference-Set Composition . . 47

6 Hypothesis Evaluation and the Desset Prior 48
6.1 The Six-Diagnostic Battery 48
6.2 Prior Elicitation and Hypothesis Priors 49
6.3 Bayes Factor Computation: $\mathcal{H}_1$ vs. $\mathcal{H}_2$ 49
6.4 The Named-Entity Alignment: A Detailed Account 50
6.5 Evaluating the Completeness of the Desset Sign-Value Assignments 50
6.6 The Bilingual Constraint as a Likelihood Ratio Test 51

7 Bayesian Synthesis 53
7.1 The Full Posterior Distribution 53
7.2 Minimum Description Length Analysis 53
7.3 Sensitivity Analysis 54
7.4 What Would Change the Conclusion 54
7.5 Posterior Predictive Checks 55
7.6 Forecasting the Posterior Under Future Corpus Growth 56
7.7 The Role of the Desset Prior in Determining the Posterior 56

8 The Proto-Elamite / Linear Elamite Relationship 58
8.1 The Structural Relationship Hypothesis 58
8.2 Sign-Form Matching Analysis 58
8.3 The Encoding-Type Discontinuity 59
8.4 Quantifying the Temporal Gap and Its Implications 59
8.5 Candidate Mechanisms for the Proto-Elamite to Linear Elamite Transition . . 60
8.6 The Elamite Cuneiform Tradition as a Ground Truth for the Phonographic Hypothesis 61

9 Synthesis and Open Problems 63
9.1 Summary of Principal Findings 63
9.2 Open Problems and Future Research Directions 64
9.3 The Elamite Writing Tradition: A Quantitative Summary 64
9.4 Implications for the History of Writing in the Ancient Near East 65
9.5 A Research Agenda for the Next Decade 66

10 Case Studies: Close Readings of Selected Inscriptions 67
10.1 The Puzur-Inshushinak Table of Nations (Sb 17488) 67
10.2 The Susa Silver Tablets: A Group Analysis 67
10.3 The Shahdad Bronze Axe Inscription 68
10.4 The Tal-i Malyan Royal Inscriptions 68
10.5 The Awan Silver Tablets: A Typological Subgroup Analysis 69

© Silvano Salvador

10.6 The Persepolis Fragments 69

11 Formal Notation and Mathematical Conventions **70**
11.1 Probability and Measure Theory Conventions 70
11.2 Matrix and Linear Algebra Conventions 70
11.3 Sign Inventory Notation 71
11.4 List of Abbreviations and Technical Symbols 71

12 Statistical Methods and Regularization **73**
12.1 Bootstrap Confidence Intervals 73
12.2 Entropy Estimation Bias Correction 73
12.3 Laplace Smoothing for Conditional Entropy 73
12.4 Maximum Likelihood Estimation for Power-Law Distributions 74
12.5 Regularization Strategy for the Conditional Entropy Estimator 74
12.6 Phase-Stratified Bootstrap 75

13 Corpus Catalogue **76**
13.1 Catalogue Organization 76

14 Sign-Level Distributional Profiles **78**
14.1 Frequency Ranking and Zipf Residuals 78

15 Computational Reproducibility **80**
15.1 Software and Data 80

16 Archaeological and Cultural Context **81**
16.1 Geographical Context: The Elamite Cultural Zone 81
16.2 The Awan and Shimashki Dynasties 82
16.3 The Economic Background: Trade Networks and the Script's Social Context . 82
16.4 The Puzur-Inshushinak Historical Record 83

17 Comparative Epigraphy **84**
17.1 The Broader Context: Third-Millennium Script Comparanda 84
17.2 Linear Elamite and the Broader Syllabic Script Family 85
17.3 The Indus-Elamite Contact Zone and Its Implications for Sign Diffusion . . . 85
17.4 The Desset Decipherment Proposal: A Complete Annotation 86

18 Extended Hypothesis Space **88**
18.1 Intermediate and Mixed Hypotheses 88

19 Thematic Glossary: Key Concepts and Technical Terms **89**
19.1 Concepts from Information Theory 89
19.2 Concepts from Epigraphy and Palaeography 89
19.3 Concepts from Bayesian Model Comparison 90

20 A Historiography of Linear Elamite Scholarship **91**
20.1 The Early Period: Discovery and Initial Despair (1900–1950) 91
20.2 The Hinz Period: Syllabary Hypotheses Without Confirmation (1950–1980) . 91
20.3 The Fallow Period (1980–2015) 91
20.4 The Desset Period and Contemporary Debate (2015–Present) 92

© Silvano Salvador

21 Robustness, Simulation, and Sensitivity Analysis **93**
21.1 Monte Carlo Resampling of the Entropy Estimates 93
21.2 Sensitivity to Sign-Reading Disagreements 93
21.3 Leave-One-Inscription-Out Cross-Validation 94

22 Numerical and Metrological Notation in the Linear Elamite Corpus **95**
22.1 The Absence of a Dedicated Numerical Sub-System 95
22.2 Implications for the Genre Composition of the Corpus 95

23 Missing-Data Imputation and Sign-Frequency Forecasting **96**
23.1 The Missing-Sign Problem 96
23.2 Forecasting the Effect of Corpus Growth 96

24 A Decision-Theoretic Framework for Evaluating Future Decipherment Claims **97**
24.1 The Asymmetry of Decipherment Error 97

25 Scribal Hands, Ductus, and Individual Variation: An Extended Analysis **98**
25.1 Methodology for Scribal-Hand Identification 98
25.2 Cluster Analysis Results 98
25.3 Implications for the Standardization Question 99
25.4 The Puzur-Inshushinak Inscriptions: A Case for Royal Scribal Supervision . . 100

26 An Agent-Based Simulation of Linear Elamite Scribal Practice **101**
26.1 Motivation: From Statistical Description to Generative Mechanism 101
26.2 Model Calibration Against Corpus Statistics 102
26.3 Simulation Results: Testing the Phonographic Hypothesis 102
26.4 Interpreting the Simulation Mismatch at Later Positions 102

27 Geographical Distribution and Spatial Epigraphy **104**
27.1 The Linear Elamite Findspot Network 104
27.2 Spatial Clustering and the Question of Script Centres 104
27.3 The Konar Sandal Assemblage and Its Significance 105
27.4 Distance-Decay of Sign-Form Standardization 106

28 A Reception History of Linear Elamite Scholarship **107**
28.1 The Discovery Period and Early Hypotheses (1900–1930) 107
28.2 The Hinz-Meriggi Period (1930–1970) 107
28.3 The Fallow Period and Methodological Introspection (1970–2015) 108
28.4 The Desset Decipherment and Its Aftermath (2015–Present) 108

29 The Formal Notation and Symbol Glossary **110**
29.1 Corpus and Sign-System Notation 110
29.2 Information-Theoretic Notation 110
29.3 Bayesian and Hypothesis-Evaluation Notation 110

30 Statistical Methods and Computational Details **112**
30.1 Entropy Estimators and Bias Correction 112
30.2 Permutation Tests 113
30.3 The Zipf-Mandelbrot Fitting Procedure 113

© Silvano Salvador

31 The Proto-Elamite / Linear Elamite Relationship Revisited: Inferential Transfer 114
31.1 The Inferential Transfer Problem . . . 114
31.2 Comparison to the Linear A / Linear B Case . . . 114
31.3 Implications for the Network Horizon Framework . . . 115

32 Synthesis Across Companion Volumes 116
32.1 A Structured Comparison Across the Companion Series . . . 116
32.2 Common Methodological Lessons . . . 116
32.3 Divergences Worth Noting . . . 117

33 An Extended Case Study: The Puzur-Inshushinak Bilinguals and the Logic of Decipherment 118
33.1 Why the Bilinguals Are Central . . . 118
33.2 The Robustness Question: How Many Anchor Errors Would Overturn the Decipherment? . . . 119

34 An Agent-Based Analysis of Hypothetical Corpus Growth Scenarios 120
34.1 Why Corpus Growth Matters . . . 120
34.2 Scenario Analysis: The Administrative Register Discovery . . . 121
34.3 Implications for Research Priorities . . . 121

References 122

Formal Models of Scribal Error and Orthographic Variation in Linear Elamite 125
34.4 Distinguishing Error from Allographic Variation . . . 125
34.5 A Confusion-Matrix Approach to Sign-Substitution Error . . . 126
34.6 Implications for the Sign-Inventory Count . . . 126
34.7 Error Versus Variant in the Konar Sandal Sub-Corpus . . . 126

A Quantitative Theory of Sign Obsolescence and Innovation in Linear Elamite 127
34.8 Motivation: Sign Dynamics in a Short-Lived Corpus . . . 127
34.9 Turnover Estimates and Their Interpretation . . . 127
34.10 Are Entering Signs Phonetically Motivated? . . . 128

A Comparative Survey of Decipherment Methodologies: Linear Elamite in Cross-Script Perspective 129
34.11 Three Historical Pathways to Decipherment . . . 129
34.12 Which Pathway Does the Desset Decipherment Follow? . . . 129
34.13 What Would Advance the Decipherment Further? . . . 130
34.14 Linear Elamite in the Landscape of Undeciphered Scripts . . . 130

Pedagogical Appendix: Worked Statistical Derivations for Non-Specialists 131
34.15 Purpose of This Appendix . . . 131
34.16 Deriving the Shannon Entropy Formula . . . 131
34.17 Computing the Bayes Factor: A Worked Example . . . 131
34.18 The Bootstrap: Intuition and Application . . . 132
34.19 Reading the Zipf-Mandelbrot Plot . . . 132

Frequently Raised Objections and Responses 133
34.20 "The Corpus Is Too Small for Reliable Entropy Estimation" . . . 133
34.21 "Bayesian Priors for the Hypotheses Are Arbitrary" . . . 133
34.22 "The Desset Decipherment Has Already Settled the Question" . . . 133

© Silvano Salvador

34.23"Quantitative Methods Cannot Capture What Philologists Already Know" . . 134
34.24"The Phonographic Hypothesis Was Already Obvious From the Sign Count" 134

Paleographic Variation and the Geometry of Sign Form in Linear Elamite 135
34.25Beyond the Sign List: Quantifying Allograph Shape Variation 135
34.26A Shape-Distance Clustering of High-Frequency Linear Elamite Signs 135
34.27Are the Bimodal Signs Site- or Hand-Specific Variants? 136
34.28Recommendation for Future Sign-List Revision 136

Climate, Political Disruption, and the End of the Linear Elamite Tradition 137
34.29The Disappearance of Linear Elamite: The Problem Stated 137
34.30Political Context: The Gutian Disruption and Elamite Reorganization 137
34.31The Climate Signal: The 4.2 kyr Event 138
34.32Why Elamite Cuneiform Rather Than a Linear Elamite Successor? 138

Coda: What This Monograph Can and Cannot Establish 140
34.33A Summary Ledger 140
34.34On the Value of Quantitative Method Without Complete Decipherment 140
34.35A Research Agenda for the Next Five Years 141

Extended Data Appendix: Supplementary Tables 142
34.36Complete Sign-Frequency Table 142
34.37Bootstrap Confidence Intervals: Full Table 144
34.38Reference Corpus Classification Coordinates 144

Extended Methodological Glossary: Formal Definitions Consolidated 145
34.39Information-Theoretic Concepts: Extended Definitions 145
34.40Bayesian and MDL Concepts: Extended Definitions 146
34.41Epigraphic and Palaeographic Concepts: Extended Definitions 146

Digitization, Data Provenance, and Reproducibility Infrastructure 148
34.42The Status of the Quantitative Material in This Monograph 148
34.43Why a Synthetic Dataset Was Used 148
34.44Generation Procedure for the Synthetic Dataset 149
34.45Data Availability Statement 149

Computational Stemmatics: Scribal Schools and Sign-Form Transmission in Linear Elamite 151
34.46The Concept of Scribal Stemmatics 151
34.47Methods: Sign-Form Distance as a Stemmatic Proxy 152
34.48Robustness: Bootstrap Stability of the Two-Cluster Partition 152
34.49Implications for the Transmission of Sign Conventions (Methodological Reading) 153

A Formal Theory of Scribal Memory: Recurrence, Long-Range Dependence, and Formulaic Sequences in Linear Elamite 154
34.50Scribal Memory as a Stochastic Process 154
34.51Hurst Exponent Estimates for the Linear Elamite Corpus 154
34.52Formulaic Sequences as the Mechanism 154
34.53Implications for Corpus-Level Entropy Estimates 155

© Silvano Salvador

Trade Networks and Foreign Contact: Epigraphic Evidence from Outside Susiana 156
34.54Linear Elamite Beyond the Core Region . . . 156
34.55Konar Sandal: A Regional Linear Elamite Tradition . . . 156
34.56The Trade Context: Konar Sandal and the Halil Rud Network . . . 157
34.57Quantitative Evidence for the Two-Region Trade Model . . . 157

Material Analysis of Inscribed Surfaces and Its Evidentiary Value 158
34.58Why Substrate Matters for Script Analysis . . . 158
34.59Legibility Profiles by Substrate Type . . . 158
34.60Metallurgical Analysis and Dating . . . 159
34.61Implications for Corpus Bias . . . 159

Typological Universals and Linear Elamite: A Cross-System Test 160
34.62Typological Universals in Writing Systems . . . 160
34.63Testing Universal U1: Inventory Size and Stroke Complexity . . . 160
34.64Testing Universal U3: Frequency and Sign Complexity . . . 160
34.65Testing Universal U4: Positional Entropy . . . 161

The Economics of Scribal Labor in the Linear Elamite Tradition 162
34.66Quantifying Scribal Labor Input . . . 162
34.67Labor Distribution Across the Corpus . . . 162
34.68Implications for Understanding Institutional Context . . . 162

Authenticity, Provenance, and Forgery Detection in the Linear Elamite Corpus 164
34.69The Provenance Problem in Ancient Epigraphy . . . 164
34.70Results: Internal Consistency of the Unprovenanced Objects . . . 164
34.71Sign-Form Analysis: Are Unprovenanced Sign Forms Consistent? . . . 165

Quantitative Iconography of Seal Impressions Associated with Linear Elamite Inscriptions 166
34.72Seals, Sealings, and the Administrative Context . . . 166
34.73Iconographic Content Distribution . . . 166
34.74Co-occurrence Analysis: Does Iconographic Category Predict Text Length? . . . 167

Quantitative Cross-Script Benchmarking: An Extended World Sample 168
34.75Beyond the Core Reference Corpus . . . 168
34.76Extended Reference Corpus: 24 Scripts . . . 168
34.77Classification in the Extended Space . . . 168
34.78The Phaistos Disc in the Extended Space . . . 169

Spatial Statistics of Inscription Find-Contexts Within Sites 170
34.79Intra-Site Spatial Distribution: A Statistical Approach . . . 170
34.80Results: Spatial Clustering at Susa . . . 170
34.81Implications for Functional Interpretation . . . 171

A Formal Information-Theoretic Comparison of the Six Undeciphered Scripts in This Companion Series 172
34.82The Companion Series as a Comparative Framework . . . 172
34.83Similarities and Differences Across the Series . . . 172
34.84What Linear Elamite's Profile Contributes to the Comparative Picture . . . 173

© Silvano Salvador

The Elamite Language and Its Phonological Structure: Implications for Script Design 174
34.85 Elamite as a Language Isolate . . . 174
34.86 The Desset Phonological Grid . . . 174
34.87 Phonological Implications for the Low-Credibility-Weight Assignments . . . 175

Extended Sign-Value Matrix: A Complete Tabulation Under All Proposed Decipherment Systems 176
34.88 Three Generations of Decipherment Proposals . . . 176
34.89 Convergence and Divergence Across Generations . . . 177

The Digital Humanities Infrastructure for Linear Elamite Research 178
34.90 The Current State of Digital Resources . . . 178
34.91 Data Standards and Interoperability . . . 178
34.92 Machine Learning Applications: Current Limitations . . . 179

Reception and Impact: Linear Elamite in Contemporary Scholarship and Public Discourse 180
34.93 The 2022 Breakthrough and Its Reception . . . 180
34.94 The Strong Claim: Encoding-Type Confirmation . . . 180
34.95 The Weaker Claim: Individual Sign-Value Assignments . . . 180
34.96 Open Questions Driving Current Research . . . 181

Network Analysis of Sign Co-occurrence: A Graph-Theoretic Perspective 182
34.97 Sign Co-occurrence as a Network . . . 182
34.98 Network-Theoretic Statistics . . . 182
34.99 Hub Signs and Their Interpretation . . . 183

Minimum Description Length Analysis of Alternative Sign-Grouping Hypotheses 184
34.100 The Sign-Grouping Problem . . . 184
34.101 Optimal Merging Under MDL . . . 184
34.102 Impact on the Entropy Estimate . . . 185

Conclusion: Linear Elamite as a Case Study in Epistemically Honest Decipherment 186
34.103 What Has Been Demonstrated . . . 186
34.104 The Broader Contribution: Method as Contribution . . . 187
34.105 A Final Reflection on the Limits of Method . . . 187

Diachronic Entropy Drift: Did the Linear Elamite Corpus Evolve Over Time? 188
34.106 Motivation . . . 188
34.107 Results . . . 188
34.108 Implications . . . 189

A Probabilistic Model of the Undiscovered Linear Elamite Corpus 190
34.109 How Much Remains to Be Found? . . . 190
34.110 Projected Inventory Growth with Corpus Expansion . . . 190
34.111 Implications for the Encoding-Type Classification . . . 191

Reflections on Uncertainty: A Methodological Postscript 192
34.112 Uncertainty as the Primary Output . . . 192
34.113 The Specific Uncertainties That Matter Most . . . 192
34.114 An Invitation to Future Researchers . . . 193

© Silvano Salvador

Supplementary Bibliography: Annotated Guide to Primary Sources **194**
34.115 Foundational Corpus Editions 194
34.116 Archaeological and Historical Context 194
34.117 Comparative Script Studies 195
34.118 Statistical and Computational Methods 195

Consolidated Index of Formal Definitions, Theorems, and Findings **196**
34.119 Purpose and Organization 196
34.120 Summary Statistics at a Glance 198

Note on Companion Volumes **200**

© Silvano Salvador

Abstract

This monograph develops a mathematically rigorous framework for the analysis of the Linear Elamite script, a corpus of approximately 43 surviving monumental and prestige-object inscriptions produced in the Elamite cultural zone of southwestern Iran during the late third millennium BCE (*c.* 2300–1800 BCE), with the majority datable to the Awan and early Shimashki dynastic periods and concentrated at Susa but attested also at Persepolis, Shahdad, and several smaller findspot clusters. The script occupies a peculiar position in the history of writing studies: it has, for nearly a century, resisted sustained decipherment despite the existence of several apparent trilingual or bilingual monuments in which Linear Elamite inscriptions are juxtaposed with Akkadian cuneiform texts of known content. A provisional decipherment proposal advanced by Desset and colleagues between 2020 and 2024, claiming syllabic phonographic values for a majority of the sign inventory, has generated substantial scholarly debate but has not yet achieved full consensus. The present monograph treats that proposal as a testable hypothesis to be evaluated alongside its principal competitors using a battery of mathematical diagnostics formally parallel to the methodology developed in the companion volumes on proto-Elamite and Linear A.

The analysis proceeds on four analytically independent axes. The first is material and epigraphic: a Bayesian assessment of the monument dating evidence, inscription-support typology, and the spatial distribution of findspots is combined with sign-form palaeographic analysis to produce a formal chronological partition of the corpus into Early, Middle, and Late Linear Elamite phases. The second axis is information-theoretic: entropy estimation, Zipf-law regression, positional-profile analysis, and model-order ceiling computation are applied to the attested sign sequences, with systematic comparison against the partly-overlapping proto-Elamite corpus, the Akkadian cuneiform texts on the bilingual monuments, and a typological reference panel of syllabic, logographic, and mixed scripts. The third axis is statistical-linguistic: conditional entropy profiles, transition-probability matrices, and positional bigram structure are interrogated to assess the degree to which the corpus is consistent with full syllabic phonography, constrained logographic notation, or hybrid encoding. The fourth axis is Bayesian model comparison: prior distributions over sign-inventory properties, Bayes factors across the principal competing hypotheses, and Minimum Description Length differentials are computed to produce a formally grounded posterior ranking of the four hypotheses.

The principal findings are as follows. The epigraphic evidence supports a three-phase chronological structure for the corpus, with sign-form variation across phases consistent with gradual palaeographic drift rather than abrupt reform. The information- theoretic profile of the Linear Elamite sign sequences is consistent with phonographic syllabic encoding at order-of-magnitude level: the unigram entropy is substantially lower than proto-Elamite, the conditional entropy profile declines in a manner qualitatively similar to attested syllabic scripts, and the Zipf exponent falls within the range observed for phonographic systems rather than logographic ones. However, the corpus size — at approximately 1,500 attested sign tokens across 43 inscriptions — is sufficiently small that the information-theoretic evidence alone does not discriminate reliably between the full syllabic hypothesis and a logosyllabic or limited-

logographic alternative. The Bayesian model comparison, incorporating the Desset phonographic assignments as structured prior evidence alongside the corpus- internal statistical diagnostics, yields posterior odds in favour of the syllabic hypothesis of approximately 7:1 over the logographic alternative, a result this monograph classifies as *substantial but not decisive* evidence. The monograph concludes by identifying the specific corpus extensions — additional inscriptions, better-quality squeezes of known monuments, and refined Akkadian-parallel alignment — that would be required to cross the threshold to decisive evidential support.

Keywords: Linear Elamite script, Susa, Awan dynasty, information theory, conditional entropy, Zipf's law, Bayesian epigraphic analysis, syllabic phonography, proto-Elamite, Akkadian cuneiform, undeciphered scripts, Desset decipherment.

Note on the Status of the Quantitative Material in This Monograph

This monograph is a *methodological* work: its purpose is to develop and demonstrate a mathematical and computational framework for evaluating decipherment hypotheses about undeciphered scripts. With the exception of explicitly cited published findings (e.g. the existence and general content of the Desset *et al.* decipherment proposal, the approximate corpus size, and the bilingual monuments), the numerical results presented here—entropy values, Zipf exponents, Bayes factors, credibility weights, cluster statistics, confidence intervals, and the accompanying token-frequency dataset—are computed on a *synthetic, algorithmically generated* illustrative dataset constructed to have plausible statistical properties for a small syllabic-scale corpus, not on a verified transcription of the actual surviving Linear Elamite inscriptions. They should be read as a worked demonstration of *how* the framework would be applied to a real corpus of this kind, not as an empirical finding about the real script or about the correctness of any real scholar's decipherment proposal. Full disclosure of the data status and generation procedure is given in the chapter *Digitization, Data Provenance, and Reproducibility Infrastructure* and in the Data Availability Statement therein.

© Silvano Salvador

Chapter 1

The Problem and Its Mathematical Structure

1.1 The Inherited Debate and Its Epistemological Deficits

The Linear Elamite script stands in an epistemological position that is simultaneously more and less tractable than the other undeciphered writing systems examined in this series of monographs. It is more tractable than proto-Elamite because it possesses what proto-Elamite conspicuously lacks: apparent bilingual monuments — chiefly the inscribed objects associated with the Awan king Puzur-Inshushinak ($fl.\ c.$ 2150 BCE) — in which Elamite inscriptions in Linear Elamite and cuneiform respectively appear on the same object, the latter in a well-understood Akkadian-period idiom. It is less tractable than Linear A because the internal redundancy of the corpus is severely limited by the corpus's small size: 43 known inscriptions, many fragmentary, yielding approximately 1,500 sign tokens after conservative quality filtering, a corpus roughly six times smaller than the Linear A record and twenty times smaller than the usable proto-Elamite non-numerical sub-corpus. These two structural facts — the existence of external anchors and the scarcity of internal statistical signal — jointly determine the mathematical strategy of this monograph.

Definition 1.1: The Epistemological Problem

The Linear Elamite decipherment problem is formally characterized by three concurrent structural features: (i) the existence of apparent bilingual monuments providing partial external anchors — semantic equivalence classes between Akkadian textual content and Linear Elamite inscriptions on the same object — without full sign-by-sign phonological assignments; (ii) a corpus sufficiently small to limit reliable statistical inference to unigram and partial bigram analyses, below the threshold for the higher-order diagnostics that yield the most discrimination in larger corpora; and (iii) a contested provisional decipherment (Desset et al. 2020–2024) that provides structured prior information amenable to Bayesian treatment but requires independent evidential evaluation before acceptance. The present monograph treats all three features as quantitative inputs to a formal analysis rather than qualitative obstacles to be discussed and set aside.

Note 1.1: The Scope of Prior Quantitative Work

The quantitative study of Linear Elamite has been substantially thinner than for proto-Elamite, reflecting the much smaller corpus. The most systematic pre-2020 treatment

is the sign list of Meriggi (1971–1974), revised and extended by Hinz and Koch (1987) and by Stolper and Tavernier (2007). The Desset et al. provisional decipherment (first published as Desset 2020 and substantially revised through 2024 with the collaboration of Kahl, Kelany, and others) represents the most comprehensive recent attempt to assign phonographic values to the majority of signs. No prior study has subjected the corpus to information-theoretic analysis in the sense developed in the present monograph; the closest antecedents are the brief entropy calculations in Rao (2010) for comparative typological purposes and the sign-frequency discussion in Mäder et al. (2017). The present work constitutes the first sustained quantitative treatment of the Linear Elamite sign corpus in the information-theoretic tradition.

1.2 The Core Mathematical Object: A Formal Description of the Corpus

Definition 1.2: The Linear Elamite Corpus $\mathcal{C}_{LE}$

The Linear Elamite corpus $\mathcal{C}_{LE}$ is defined as the ordered set of sign-token sequences attested on the $N_{ins} = 43$ inscriptions and inscription fragments accepted by this monograph after a quality filter requiring (i) at least three consecutive readable signs; (ii) publication in a scholarly edition with photographic or squeeze documentation; and (iii) a findspot attribution to the geographical Elamite zone. The resulting working corpus contains approximately $N_{LE} \approx 1{,}480$ sign tokens over $|\Sigma_{LE}| \approx 71$ distinct sign types, following the sign-list conventions of Desset (2020) as revised through 2023. The corpus is partitioned into:

- A *parallel-text sub-corpus* $\mathcal{P}$ of ≈ 340 tokens from inscriptions on objects also bearing Akkadian cuneiform texts of identifiable content, providing the external anchor used in Chapter 6;
- An *independent sub-corpus* $\mathcal{I}$ of $\approx 1{,}140$ tokens from inscriptions with no usable external parallel, providing the primary internal-statistical dataset for Chapters 3–5.

The full corpus $\mathcal{C}_{LE} = \mathcal{P} \cup \mathcal{I}$ is used for sign-inventory and frequency analyses; the two sub-corpora are analysed separately where the distinction is analytically relevant.

Definition 1.3: The Sign Inventory Σ_{LE}

The Linear Elamite sign inventory is defined as the set of graphically distinct sign types attested in $\mathcal{C}_{LE}$, following the sign-list conventions of Desset (2020). The working inventory size is $|\Sigma_{LE}| = 71$ distinct types, of which $|\Sigma^*_{LE}| = 53$ are attested at least twice in $\mathcal{C}_{LE}$ and constitute the *reduced inventory* used for frequency-distribution analyses. Hapax legomena ($n = 18$ types attested exactly once) are excluded from distribution analyses but retained for inventory-size calculations and discussed in Section 3.5 in the context of the corpus-size correction for Zipf-law estimation. The inventory of 71 signs is substantially smaller than the proto-Elamite inventory ($|\Sigma_{PE}| \approx 1{,}000$) and is in the range expected for syllabic scripts: the Cypro-Minoan inventory has 56 signs, Linear A has 55 syllabograms, the Sumero-Akkadian syllabary in the Ur III period has approximately 50–80 actively used syllabic values. This inventory-size observation constitutes

© Silvano Salvador

the first of the six diagnostics developed in Chapter 6.

Proposition 1.1: Corpus-Size Constraints on Statistical Inference

At $N_{LE} \approx 1{,}480$ tokens over an inventory of $|\Sigma^*_{LE}| = 53$ types, the Linear Elamite corpus supports reliable unigram entropy estimation (coverage ratio $\rho_0 = N_{LE}/|\Sigma^*_{LE}| \approx 27.9$, well above the adequacy threshold of $\rho_0 \geq 5$) and partial first-order conditional entropy estimation ($\rho_1 = N_{LE}/|\Sigma^*_{LE}|^2 \approx 0.53$, below the adequacy threshold of $\rho_1 \geq 1$ for complete bigram coverage, requiring regularization). Reliable unregularized bigram analysis is not supported for the full inventory, though it becomes feasible for the high-frequency sub-inventory of the 20 most common signs ($\rho_1^{(20)} = N_{LE}/20^2 = 3.7$). All analyses are conducted at the statistically appropriate order with regularization strategies specified in Appendix B; results depending on bigram statistics are flagged accordingly and their sensitivity to regularization choice is reported.

1.3 The Four Analytical Axes and Their Relationship

Definition 1.4: The Four Analytical Axes

This monograph develops its analysis along four independent axes, each producing evidence bearing on the hypothesis-comparison problem of Section 1.4:

(i) **Material and epigraphic:** Bayesian assessment of monument dating, inscription-support typology, spatial distribution of findspots, and palaeographic sign-form drift analysis, producing a formal three-phase internal chronology of the corpus (Chapter 2);

(ii) **Information-theoretic:** Shannon entropy estimation, Zipf-law regression, positional entropy profiling, model-order ceiling computation, and the phonographic consistency diagnostic developed in Chapter 3;

(iii) **Statistical-linguistic:** conditional entropy profiles, transition- probability matrices, and positional bigram structure interrogated against the typological reference panel of Chapter 5; cross-entropy against attested syllabic scripts (Cypro-Minoan, Linear B, Carian); and transfer entropy across the chronological phases;

(iv) **Bayesian model comparison:** Prior elicitation incorporating the Desset phonographic assignments, Bayes factor computation across the four hypotheses, MDL differential calculation, and posterior ranking (Chapters 6–7).

The four axes are designed to be methodologically independent: each draws on different aspects of the corpus and makes different distributional assumptions. Their joint conclusions are more reliable than any single axis alone, and their pattern of agreement and disagreement is itself an informative diagnostic about the robustness of the overall inference.

© Silvano Salvador

1.4 Formalizing the Hypothesis Space

Definition 1.5: The Four Principal Hypotheses $\mathcal{H}_{LE}$

The hypothesis space for the Linear Elamite sign system is:

$\mathcal{H}_1$: Full syllabic phonography (the Desset hypothesis);
$\mathcal{H}_2$: Logosyllabic encoding (mixed logo-syllabic, Elamite language);
$\mathcal{H}_3$: Constrained administrative logography (semasiographic);
$\mathcal{H}_4$: Prestige display script without full linguistic encoding.

These four hypotheses form a partially ordered set under the relation of increasing linguistic encoding density, from $\mathcal{H}_4$ (least) to $\mathcal{H}_1$ (most). They are mutually exclusive as stated. The formal evaluation of Chapters 6–7 computes posterior probabilities for each; intermediate mixed hypotheses (e.g., $\mathcal{H}_1$ restricted to a limited phonographic sub-set) are addressed in the extended hypothesis space of Appendix H.

Definition 1.6: The Syllabic Phonography Hypothesis in Detail

$\mathcal{H}_1$ (full syllabic phonography) is defined formally as the hypothesis that each non-numerical Linear Elamite sign type encodes a CV, V, or VC syllabic unit of a spoken language — identified as Elamite, or a closely related ancestral form thereof, by the Desset proposal — such that sign sequences on each inscription constitute a linear phonographic transcription of an Elamite utterance, readable in principle by a competent speaker of that language with appropriate sign-value knowledge. This hypothesis predicts: (i) a sign inventory in the range 40–100 typical of syllabic scripts; (ii) a Zipf exponent α in the range $[0.9, 1.4]$ observed for syllabic scripts as opposed to $[0.6, 0.9]$ for logographic systems; (iii) a conditional entropy profile with $H_1/H_0 < 0.75$ (substantial reduction in uncertainty given the preceding sign, due to morphological and phonological constraints of the Elamite language); and (iv) a significant positive correlation between inscription length and inscription content complexity as approximated by the Akkadian parallels. Predictions (i)–(iv) are quantified and tested in Chapters 3–5.

Thesis 1.1: The Central Thesis of This Monograph

The principal finding of this monograph, established by the convergent evidence of all four analytical axes, is that the corpus-internal statistical evidence is *consistent with but does not compel* $\mathcal{H}_1$ (full syllabic phonography). The inventory size, Zipf exponent, and conditional entropy ratio all fall within the ranges predicted by $\mathcal{H}_1$, and the Bayesian synthesis yields posterior odds in favour of $\mathcal{H}_1$ over $\mathcal{H}_2$ of approximately 7:1. However, three independent diagnostic limitations prevent a decisive verdict: the corpus size is insufficient to exclude $\mathcal{H}_2$ at a conventionally decisive Bayes factor (BF ≥ 100); the bilingual-monument alignments are insufficient for direct sign-value verification of more than a handful of proposed assignments; and the Desset phonographic values, while internally consistent for the signs they cover, are not independently validated by a second decipherment pathway. This monograph therefore concludes with a formal specification of the evidential requirements for a decisive outcome, targeted at the future epigraphic and archaeological programme.

© Silvano Salvador

1.5 Relationship to the Companion Volumes

Note 1.2: The Series Framework

This monograph is the fourth in a series applying the same mathematical framework to undeciphered scripts. The first volume (Silvano Salvador, *The Voynich Manuscript: A Mathematical Analysis*) established the methodology; the second (Silvano Salvador, *Linear A: A Mathematical Analysis*) applied it to a script whose natural-language status is positively established; the third (Silvano Salvador, *The Proto-Elamite Script: A Mathematical Analysis*) analysed the large but statistically intractable proto-Elamite corpus. The present volume completes the Elamite arc of the series: Linear Elamite was long conjectured to be a descendant or lateral development from proto-Elamite, though the structural relationship between the two scripts — if any — remains contested; the comparative analysis of Chapter 8 addresses this question quantitatively for the first time in a formally explicit framework. The four volumes together span the full range of epistemological positions available for undeciphered scripts: near-certain meaninglessness (Voynich), confirmed linguistic encoding with unknown language (Linear A), confirmed non-phonographic encoding (proto-Elamite), and provisionally phonographic encoding under active scholarly debate (Linear Elamite).

1.6 Structure of the Volume

The volume proceeds through material and epigraphic chronology (Chapter 2), information-theoretic profiling (Chapter 3), bigram and positional analysis (Chapter 4), reference-corpus classification (Chapter 5), hypothesis evaluation and the Desset prior (Chapter 6), Bayesian synthesis (Chapter 7), and a dedicated comparative analysis of the proto-Elamite/Linear Elamite relationship (Chapter 8), before closing with synthesis and open problems (Chapter 9) and a set of close-reading case studies of individual inscriptions (Chapter 10). An appendix collects formal notation (Appendix A), statistical methods and regularization strategies (Appendix B), a corpus catalogue with full quality assessments (Appendix C), sign-level distributional profiles (Appendix D), computational reproducibility (Appendix E), the archaeological and cultural context (Appendix F), comparative epigraphy (Appendix G), and the extended hypothesis space (Appendix H).

1.7 The Three-Century Window: Historical and Archaeological Context

Linear Elamite inscriptions are datable, on epigraphic and historical grounds, to a window of roughly three centuries from approximately 2300 BCE to approximately 2000 BCE, with possible extensions to 1800 BCE for the latest stylistically assignable monuments. This window overlaps substantially with the Akkadian Empire period in Mesopotamia and with the documented Awan dynastic period in Elam, both of which provide the external historical anchors for the most informative bilingual monuments.

Definition 1.7: The Three Chronological Phases of the Corpus

On the basis of the palaeographic, material, and historical analysis of Chapter 2, the Linear Elamite corpus is partitioned into three phases:

1. **LE-I (*c.* 2300–2150 BCE):** Pre-Puzur-Inshushinak inscriptions from the Awan

© Silvano Salvador

period, attested on silver vases, silver tablets, and stone monuments. Approximately 8 inscriptions with 310 tokens.

2. **LE-II (*c.* 2150–2000 BCE):** The Puzur-Inshushinak corpus and contemporaneous material, including the bilingual and trilingual table-of-nations inscriptions that provide the primary external anchors. Approximately 22 inscriptions with 780 tokens.
3. **LE-III (*c.* 2000–1800 BCE):** Post-Akkadian, Shimashki-period and later material. Approximately 13 inscriptions with 390 tokens. Palaeographic forms are more regular and slightly more cursive than LE-I/II.

This partition is formal rather than rigid: 6 inscriptions cannot be securely assigned to a single phase and are assigned probabilistically using the Bayesian model of Chapter 2.

Proposition 1.2: The Bilingual Monument Constraint

The Puzur-Inshushinak bilingual monuments — chiefly the "Table of Nations" (Louvre Sb 17488), the statue inscription (Sb 88), and the brick inscriptions (Sb 12) — provide the following formal constraint: any assignment of phonographic values to Linear Elamite signs must yield, when applied to the Linear Elamite side of those monuments, texts that are semantically compatible with the Akkadian text on the same object. This constraint does not require word-for-word equivalence (since the two versions may be stylistically distinct renderings of the same content) but does require thematic equivalence at the level of named referents (personal names, titles, divine names) and expressed propositions (conquest, dedication, building). The Desset decipherment proposal is evaluated in Chapter 6 against this constraint using a formal semantic-compatibility score derived from the Akkadian parallel.

1.8 The Writing-System Typology: A Formal Taxonomy

Before proceeding to corpus analysis, it is necessary to fix a formal taxonomy of writing-system types that will be used consistently throughout this monograph. The taxonomy adopted here follows the information-theoretic tradition of Sproat (2010) and Daniels and Bright (1996), formalized in the framework of Appendix A.

Definition 1.8: Writing System Typological Classes

A writing system $\mathcal{W}$ over alphabet Σ is assigned to one of four typological classes on the basis of the conditional mutual information between sign tokens and linguistic units:

$\mathcal{W} \in$ **Phonographic** if there exists a surjective mapping $\phi : \Sigma \to \Sigma_{phon}$ to phonological units (phonemes, syllables, or moraic segments) such that $I(\phi(S); L) > I(S; L) - \varepsilon$ for small $\varepsilon > 0$, where L denotes the linguistic content of a text and S its sign sequence. Phonographic systems include alphabets, syllabaries, and abjads.

$\mathcal{W} \in$ **Logographic** if there exists a surjective mapping $\psi : \Sigma \to \Sigma_{lex}$ to lexical units (morphemes or words) such that $I(\psi(S); L) > I(S; L) - \varepsilon$. Logographic systems include Sumerian logograms, Chinese characters in their lexical use, and the logographic component of logosyllabic mixed systems.

$\mathcal{W} \in$ **Semasiographic** if the sign system conveys meaning directly without systematic phonological or lexical correspondence: the signs encode semantic content but not the

© Silvano Salvador

phonological or lexical form of a specific natural language. Proto-Elamite is the canonical case.

$\mathcal{W} \in$ **Display** if the sign system's primary function is symbolic or aesthetic rather than communicative in the strict linguistic sense: it identifies its bearer as a member of an elite, marks a prestige object, or performs a ritual function, but does not systematically encode a linguistic message. Many partially understood prestige scripts may fall into this class.

Note 1.3: The Logosyllabic Mixed Class

Many attested writing systems of the ancient Near East and Eastern Mediterranean are *logosyllabic*: they combine logographic signs (encoding morphemes or words directly) with syllabic phonographic signs (encoding phonological units), with the proportion of each type varying by text genre, period, and scribal tradition. The Sumero-Akkadian cuneiform system at its mature stage, Egyptian hieroglyphic writing, Luwian hieroglyphic, and Maya writing are all logosyllabic in this sense. The hypothesis $\mathcal{H}_2$ of this monograph (Definition 1.5) assigns Linear Elamite to this class. The information-theoretic distinction between a predominantly syllabic logosyllabic system ($\mathcal{H}_2$) and a pure syllabic system ($\mathcal{H}_1$) is subtle and is the central discriminative challenge of this monograph, since both predict similar entropy profiles at the sign-sequence level.

Proposition 1.3: The Typological Classes Form a Monotone Chain Under Linguistic Encoding Density

Define the *linguistic encoding density* $\delta(\mathcal{W})$ of a writing system $\mathcal{W}$ as the mutual information $I(S;L)/H(L)$ between the sign sequence and the linguistic content, normalized by the entropy of the linguistic content. The four typological classes satisfy $\delta(\textbf{Phonographic}) \geq \delta(\textbf{Logographic}) \approx \delta(\textbf{Logosyllabic}) \geq \delta(\textbf{Semasiographic}) \geq \delta(\textbf{Display})$ as a chain in the partial order on classes induced by the $\leq$ relation on the interval $[0,1]$. The inequality $\delta(\textbf{Phonographic}) \geq \delta(\textbf{Logographic})$ holds because phonographic encoding transmits the full acoustic form of a linguistic message (including morphophonological distinctions not represented in lexical encoding), while logographic encoding transmits only the lexical/morphological form. The inequality $\delta(\textbf{Semasiographic}) \geq \delta(\textbf{Display})$ holds because semasiographic systems, though not encoding a specific natural language, do transmit content that can be systematically decoded (administrative records, numerical data); display scripts transmit identity or prestige markers without decodable propositional content.

1.9 Quantitative Palaeography: Methodology and Scope

A central methodological innovation of this monograph, paralleling the approach developed in the companion proto-Elamite volume, is the application of quantitative palaeographic methods to the formal study of sign-form variation across the corpus. Quantitative palaeography treats the visual forms of sign tokens as data points in a high-dimensional feature space, enabling statistical tests of chronological variation, scribal-hand differentiation, and sign-type boundaries that are not available to traditional qualitative palaeographic analysis.

© Silvano Salvador

Definition 1.9: The Palaeographic Feature Space $\mathcal{F}_{pal}$

The palaeographic feature space $\mathcal{F}_{pal} = \mathbb{R}^{d_\phi}$ with $d_\phi = 8$ (as defined in Definition 2.2) is endowed with the standard Euclidean metric. Sign tokens are represented as points in $\mathcal{F}_{pal}$ via the feature extraction procedure described in Appendix B.7, which operates on digitized photographic or squeeze tracings of the inscriptions. The feature space has the following geometric interpretation: the eight coordinate axes correspond to stroke count (integer-valued), mean curvature, aspect ratio, angular complexity, closure ratio, vertical position, component count, and incision depth, with the latter four continuous-valued. Within-inscription variation in sign-form (due to different execution instances of the same sign type) defines an intra-type cluster in $\mathcal{F}_{pal}$; inter-type variation defines the between-type structure. The palaeographic chronological analysis of Chapter 2 uses the temporal structure of these clusters to infer phase-conditional sign-form distributions.

Proposition 1.4: Sign-Type Boundaries Are Well-Separated in the Palaeographic Feature Space

For the 21 sign types attested in all three phases (the set $\mathcal{S}_{\text{multi}}$ of Theorem 2.1), the mean ratio of between-type to within-type Euclidean distance in $\mathcal{F}_{pal}$ is:

$$\bar{r}_{bw} = \frac{1}{|\mathcal{S}_{\text{multi}}|(|\mathcal{S}_{\text{multi}}|-1)} \sum_{s \neq s'} \frac{d(c_s, c_{s'})}{\frac{1}{2}(r_s + r_{s'})} = 4.7,$$

where c_s is the centroid of the token cluster for sign s in $\mathcal{F}_{pal}$ and r_s is the mean within-cluster radius. A ratio of 4.7 indicates that sign-type centroids are on average 4.7 times farther apart than the within-type spread, confirming that the palaeographic feature space provides well-separated sign-type clusters and that the feature vectors of Definition 2.2 are discriminative for sign-type identity as well as chronological drift.

1.10 The Elamite Language: Phonological and Morphological Background

The phonographic hypothesis ($\mathcal{H}_1$) predicts that Linear Elamite encodes the Elamite language. It is therefore essential to fix the relevant background on the Elamite language's phonological structure, morphological patterns, and lexical properties, since these determine the quantitative predictions of $\mathcal{H}_1$ for the statistical diagnostics of Chapters 3–5.

Definition 1.10: The Elamite Phonological System as Reconstructed from Later Cuneiform Attestations

The Elamite language, attested in cuneiform from approximately 2400 BCE (the Awan period, contemporaneous with the Linear Elamite corpus) through 360 BCE (the Achaemenid administrative texts from Persepolis), has the following phonological properties relevant to the syllabic hypothesis:

- A consonant inventory of approximately 20–24 phonemes, including stops (/p, b, t, d, k, g/), fricatives (/f, s, š, h/), nasals (/m, n/), liquids (/r, l/), and the semivowels (/w, j/), with the exact inventory uncertain due to under-differentiation in the cuneiform orthography;

© Silvano Salvador

- A vowel system of four to five vowels (/a, e, i, u/, with /o/ marginal), with vowel length phonologically contrastive in at least some periods;
- A syllable structure preference for open syllables (CV) over closed syllables (CVC or VC), consistent with the predominance of CV-type syllabic signs in the Desset phonographic assignments;
- An agglutinative morphology with suffixing nominal and verbal paradigms, producing characteristic suffix sequences that would appear as systematic sign-following patterns in a phonographic transcription.

These properties generate specific quantitative predictions for the conditional entropy profile (Section 3.3) and the bigram structure (Chapter 4) of any phonographic transcription of Elamite, predictions that are tested against the observed corpus.

Note 1.4: Elamite as a Language Isolate

Elamite is a language isolate: it has no demonstrated genealogical relationship to any other attested language family. This has two important implications for the present analysis. First, there is no closely related reference language from which the phonological or morphological predictions of $\mathcal{H}_1$ can be refined; all quantitative predictions must be derived from the cuneiform Elamite attestations themselves or from typological universals. Second, the absence of a known linguistic relative means that comparative-linguistic evidence cannot serve as an independent check on the Desset phonographic assignments, unlike the situation for (e.g.) Linear B, where knowledge of Mycenaean Greek provided an independent pathway for decipherment verification. This limitation is a structural feature of the epistemological problem and is reflected in the credibility weights assigned to the Desset prior in Section 6.1.

© Silvano Salvador

Chapter 2

Material Evidence and Epigraphic Chronology

2.1 Inscription Supports and Their Chronological Significance

The physical support of a Linear Elamite inscription carries substantial chronological information independent of its textual content. The known corpus spans four primary support types: silver objects (tablets, vases, and containers), stone monuments (stelae, statues, and orthostats), fired brick, and copper alloy objects. Each support type has a distinct distribution over the three phases defined in Chapter 1, and the typological associations of the support with known administrative and commemorative practices of the Awan and Shimashki periods constrain the dating inference.

Definition 2.1: The Inscription-Support Typology $\mathcal{T}_{LE}$

The inscription-support typology for the Linear Elamite corpus is defined as the set $\mathcal{T}_{LE} = \{T_1, T_2, T_3, T_4, T_5\}$ where:

- T_1: Silver tablets (flat rectangular objects with inscribed surfaces);
- T_2: Silver or gold vases and containers with inscribed bands;
- T_3: Stone monuments (statues, stelae, architectural orthostats);
- T_4: Fired brick (inscribed during or after firing);
- T_5: Copper and bronze objects (axes, vessels, pins).

The support type is recorded for each inscription in the corpus catalogue (Appendix C) and enters the chronological model of Section 2.3 as a discrete covariate with phase-conditional prior distributions derived from the Awan/Shimashki archaeological literature.

Note 2.1: The Susa Silver Tablets

The group of silver tablets associated with the Awan period, most of which entered collections in the early twentieth century without secure excavation provenance (the majority appearing on the antiquities market between 1900 and 1930 and attributed to Susa on stylistic and chemical grounds), constitutes the largest single cluster of Linear Elamite inscriptions. Their lack of secure archaeological context is a significant limitation: the Bayesian chronological model of Section 2.3 assigns them to LE-I with high probability ($P(\text{LE-I} \mid \text{silver tablet, stylistic profile}) \approx 0.82$) on the basis of sign-form analysis and the silver tablet typology's association with Awan-period prestige econ-

omy, but this assignment carries substantially more uncertainty than for the Puzur-Inshushinak corpus. The sensitivity of the main statistical results to the phase assignment of this group is analysed in Appendix B.4.

2.2 Palaeographic Sign-Form Analysis

The graphical form of individual sign types varies across the Linear Elamite corpus in a patterned way that carries chronological information independently of the inscription-support typology. Specifically, several high-frequency signs exhibit a systematic drift from more angular, stroked forms in LE-I to more rounded, cursive forms in LE-III, a trend consistent with the general palaeographic evolution observed in other ancient writing systems over comparable timescales.

Definition 2.2: The Palaeographic Feature Vector $\phi(s, I)$

For each sign token of type $s \in \Sigma_{LE}$ on inscription I, the palaeographic feature vector $\phi(s, I) \in \mathbb{R}^{d_\phi}$ is defined as a $d_\phi = 8$-dimensional descriptor capturing:

(1) Stroke count (integer);
(2) Mean stroke curvature (in radians per unit stroke length, from digitized squeeze tracings);
(3) Aspect ratio of the bounding box;
(4) Internal angular complexity (standard deviation of interior angle distribution);
(5) Closure ratio (fraction of strokes forming closed loops);
(6) Relative vertical position within the inscription line;
(7) Internal component count (number of disconnected stroke groups);
(8) Incision depth ratio (qualitative: 1 = shallow, 2 = medium, 3 = deep, from published description).

The mean feature vector $\bar{\phi}(s)$ over all tokens of sign s in sub-corpus $\mathcal{C}_{LE}^{(j)}$ (phase $j \in \{\text{I, II, III}\}$) is used to define the phase-specific sign form, and the drift metric $\Delta(s) = \|\bar{\phi}^{(III)}(s) - \bar{\phi}^{(I)}(s)\|_2$ measures the magnitude of palaeographic change over the three phases.

Theorem 2.1: Palaeographic Drift Is Statistically Significant

Let $\Delta_{\text{obs}} = \frac{1}{|\mathcal{S}_{\text{multi}}|} \sum_{s \in \mathcal{S}_{\text{multi}}} \Delta(s)$ denote the mean drift magnitude over the set $\mathcal{S}_{\text{multi}}$ of signs attested in all three phases ($|\mathcal{S}_{\text{multi}}| = 21$). Under the null hypothesis $\mathcal{H}_0^{\text{pal}}$ that sign-form variation is temporally uniform (no systematic drift), Δ_{obs} is distributed as a function of the within-phase variation. A permutation test (10,000 permutations of phase labels across inscriptions) yields an empirical p-value of $p_{\text{pal}} < 0.003$ for the observed Δ_{obs} relative to the permutation distribution, rejecting $\mathcal{H}_0^{\text{pal}}$ at the $\alpha = 0.01$ level. The direction of drift (toward lower stroke count, higher curvature, and lower angular complexity from LE-I to LE-III) is consistent with the cursivization trajectory observed in proto-cuneiform's evolution to Sumerian cuneiform and in Linear A's relationship to the later Cypro-Minoan tradition.

© Silvano Salvador

2.3 Bayesian Chronological Model

The Bayesian chronological model integrates the inscription-support typology ($\mathcal{T}_{LE}$), palaeographic drift measurements (ϕ), stratigraphic associations (where available), and historical anchors (the Puzur-Inshushinak dating) into a joint probabilistic model for phase assignment.

Definition 2.3: The Bayesian Phase-Assignment Model

Let $Z_I \in \{1, 2, 3\}$ denote the phase assignment of inscription I, and let $X_I = (T_I, \bar{\phi}_I, A_I, H_I)$ denote the observed covariates (support type, mean sign-form vector, archaeological association indicator, and historical anchor indicator). The Bayesian model is:

$$\begin{aligned} Z_I &\sim \text{Categorical}(\pi_1, \pi_2, \pi_3), \\ X_I \mid Z_I = j &\sim p_j(X_I \mid \theta_j), \\ (\pi_1, \pi_2, \pi_3) &\sim \text{Dirichlet}(\alpha_0), \\ \theta_j &\sim p(\theta_j), \end{aligned}$$

where $p_j(X_I \mid \theta_j)$ factorizes over the four covariate types under a conditional independence assumption given phase. The prior Dirichlet(α_0) with $\alpha_0 = (3, 5, 3)$ encodes the prior belief that LE-II is the most densely attested phase (consistent with the Puzur-Inshushinak corpus dominating the total inscription count). Posterior phase probabilities $P(Z_I = j \mid X_I)$ are computed by MCMC (10,000 samples, 2,000 burn-in, convergence verified by $\hat{R} < 1.01$ for all parameters) and are reported in Appendix C for each inscription.

Proposition 2.1: Chronological Stability of the Main Statistical Results

The main information-theoretic diagnostics of Chapter 3 — unigram entropy H_0, Zipf exponent α, and conditional entropy ratio H_1/H_0 — are computed for the full corpus $\mathcal{C}_{LE}$ and separately for the three phase-stratified sub-corpora $\mathcal{C}_{LE}^{(j)}$. The within-phase estimates are consistent with the full-corpus estimates at the level of estimation uncertainty: there is no statistically significant variation in H_0 or H_1/H_0 across phases ($p > 0.2$ for both, permutation test), indicating that the statistical profile is stable over the three centuries spanned by the corpus. The Zipf exponent α shows marginal variation ($\hat{\alpha}^{(I)} = 1.12$, $\hat{\alpha}^{(II)} = 1.09$, $\hat{\alpha}^{(III)} = 1.18$, with overlapping 95% bootstrap confidence intervals) consistent with a single underlying generative process across phases.

2.4 The Bilingual Monuments: A Formal Alignment Framework

The bilingual inscriptions of Puzur-Inshushinak provide the primary external constraint on any decipherment proposal, and their formal analysis requires an alignment framework that quantifies the semantic compatibility between the Linear Elamite text and its Akkadian counterpart without presupposing phonographic values for any Linear Elamite signs.

© Silvano Salvador

Definition 2.4: The Semantic Compatibility Score ρ_{sem}

For a bilingual inscription pair (L, A) with Linear Elamite text $L = (s_1, \ldots, s_n)$ and Akkadian text A of known content, the semantic compatibility score $\rho_{sem}(\phi_{LE}, L, A)$ for a proposed phonographic value assignment $\phi_{LE} : \Sigma_{LE} \to \Sigma_{syl}$ (mapping sign types to syllabic values) is defined as:

$$\rho_{sem}(\phi_{LE}, L, A) = \frac{1}{|E_A|} \sum_{e \in E_A} \mathbf{1}[\phi_{LE}(L) \text{ contains a phonographic form compatible with } e],$$

where E_A is the set of named entities (personal names, divine names, titles, place names) and key propositions recoverable from the Akkadian text A. The score $\rho_{sem} \in [0, 1]$ measures the fraction of expected named-entity and propositional content recoverable from the Linear Elamite side under ϕ_{LE}; a score of 1 indicates perfect compatibility and a score of 0 indicates no recoverable matching content. The Desset (2020–2023) assignment yields $\rho_{sem} \approx 0.71$ for the Puzur-Inshushinak bilingual corpus (see Chapter 6 for detailed computation).

Note 2.2: Limitations of the Bilingual Alignment

The semantic compatibility framework of Definition 2.4 is constrained by three practical limitations. First, the Akkadian texts on the Puzur-Inshushinak monuments are not literal translations of the Linear Elamite texts: they are royal inscriptions in a distinct scribal tradition, potentially expressing related content in a thematically parallel but structurally different form, so a ρ_{sem} of 1.0 may not be achievable even under a correct decipherment. Second, several of the named entities expected in the Akkadian text — in particular the Elamite divine names and the name of Puzur-Inshushinak's father — have uncertain phonological forms in the Elamite language, introducing ambiguity in the matching criterion. Third, the Linear Elamite side of the most important monument (Sb 17488) is partially damaged, reducing the effective token count available for alignment. These limitations are propagated into the uncertainty bounds on ρ_{sem} in Chapter 6.

2.5 Spatial Distribution and the Findspot Network

Definition 2.5: The Inscription Findspot Network $\mathcal{N}_{FS}$

The findspot network $\mathcal{N}_{FS} = (V_f, E_f, w_f)$ has vertex set V_f equal to the 12 distinct findspot locations (including "market provenance" as a single pseudo-vertex) at which Linear Elamite inscriptions are attested. An edge $(i, j) \in E_f$ connects two findspots if at least one sign type in Σ^*_{LE} is attested at both locations. Edge weight w_f^{ij} is the Jaccard coefficient of the sign-type sets at locations i and j: $w_f^{ij} = |\Sigma_i \cap \Sigma_j| / |\Sigma_i \cup \Sigma_j|$, where Σ_k is the set of sign types attested at findspot k. The principal findspot is Susa (17 inscriptions, $|\Sigma_{Susa}| = 58$ distinct types), followed by Persepolis/Tall-e Malyan (6 inscriptions), Shahdad (4 inscriptions), and other locations (1–2 each). The high Jaccard similarity between Susa and Persepolis ($w_f^{SP} = 0.61$) and the lower similarity between Susa and Shahdad ($w_f^{SS} = 0.38$) suggest that Shahdad either developed or preserved a partially distinct local sign repertoire.

© Silvano Salvador

2.6 Inscriptions on Metal: Silver, Gold, and Copper Objects

The Linear Elamite inscriptions on metal objects constitute a materially and iconographically distinctive sub-corpus that requires separate treatment from the stone monument inscriptions. Metal-object inscriptions are typically shorter, more formulaic, and less likely to contain the extended narratives that would provide the most informative statistical signal; however, their physical properties preserve information about the workshop contexts and economic circuits in which they were produced.

Definition 2.6: The Metal-Object Sub-Corpus $\mathcal{M}_{LE}$

The metal-object sub-corpus $\mathcal{M}_{LE} \subset \mathcal{C}_{LE}$ consists of all 23 Linear Elamite inscriptions on silver tablets, silver vases, bronze axes, bronze vessels, bronze pins, and copper alloy objects, yielding approximately 580 sign tokens. The sub-corpus has the following distributional properties:

- Mean inscription length: $\bar{n}_M = 25.2$ tokens (standard deviation: 12.8), substantially shorter than the stone monument sub-corpus mean of 57.1 tokens;
- Mean distinct sign types per inscription: $\bar{k}_M = 14.8$;
- Sign-type coverage relative to Σ^*_{LE}: 47 of 53 types (88.7%) are attested in at least one metal-object inscription;
- Initial-position concentration: the three most frequent initial-position signs account for 61% of all inscription-initial positions in $\mathcal{M}_{LE}$, compared to 48% in the stone monument sub-corpus, consistent with more formulaic opening conventions.

Proposition 2.2: The Metal-Object Sub-Corpus Shows Stronger Formulaic Concentration Than the Stone Monument Sub-Corpus

The positional entropy $H_0(1)$ of the initial position (Definition 3.5) for the metal-object sub-corpus is $H_0^{M}(1) = 2.91$ bits, compared to $H_0^{Stn}(1) = 3.82$ bits for the stone monument sub-corpus and the full-corpus value of 3.42 bits. This difference is statistically significant under a permutation test ($p < 0.01$). The finding is consistent with a scribal convention in which metal-object inscriptions begin with a more restricted set of dedicatory or ownership formulae than stone monuments, which may contain royal titulary opening sequences of greater variety. Under $\mathcal{H}_1$ (syllabic phonography), this pattern reflects the formulaic register of dedicatory Elamite prose; under $\mathcal{H}_3$ (semasiographic logography), it reflects a smaller set of relevant administrative categories for metal objects. Both explanations are evaluated in Chapter 6.

2.7 Scribal Workshop Evidence and the Question of Standardization

A critical question for any corpus of administrative or prestige inscriptions is whether the sign forms and sequence conventions observed in the corpus reflect a single scribal tradition or multiple, regionally or temporally distinct traditions. The Linear Elamite corpus, spanning three phases and multiple findspot locations, provides evidence for both scribal continuity and local variation.

© Silvano Salvador

Definition 2.7: The Scribal Workshop Similarity Index ω_{IJ}

For two inscriptions $I, J \in \mathcal{C}_{LE}$, the scribal workshop similarity index ω_{IJ} is defined as the weighted combination:

$$\omega_{IJ} = \lambda_1 \cdot \text{JSD}(p_I \| p_J)^{-1} + \lambda_2 \cdot \sigma_\phi(I, J) + \lambda_3 \cdot \delta_T(I, J),$$

where $\text{JSD}(p_I \| p_J)^{-1}$ is the inverse Jensen-Shannon divergence between the two inscription's unigram distributions (high when distributions are similar), $\sigma_\phi(I, J) = \cos(\bar{\phi}_I, \bar{\phi}_J)$ is the cosine similarity between the mean palaeographic feature vectors of the two inscriptions, $\delta_T(I, J) = \mathbf{1}[T_I = T_J]$ is the support-type match indicator, and $(\lambda_1, \lambda_2, \lambda_3) = (0.4, 0.4, 0.2)$ are empirically calibrated weights. High ω_{IJ} indicates that inscriptions I and J are likely to have been produced in the same workshop or scribal training tradition.

Theorem 2.2: The Susa Core Forms a Statistically Coherent Scribal Tradition

Hierarchical agglomerative clustering of the 43 inscriptions in $\mathcal{C}_{LE}$ using $1 - \omega_{IJ}$ as the dissimilarity measure (average linkage) yields a primary cluster of 28 inscriptions from Susa spanning all three phases, with a mean within-cluster $\omega_{IJ} = 0.71$. The Persepolis fragments, Tal-i Malyan inscriptions, and Shahdad inscriptions form secondary clusters with lower mean internal ω values (0.58, 0.54, and 0.49 respectively). The silhouette coefficient for the Susa primary cluster (0.44) is significantly above the null expectation of 0 under random clustering ($p < 0.001$, permutation test), confirming that the Susa inscriptions share a statistically coherent scribal tradition distinct from the peripheral sites. This finding supports the interpretation of Susa as the primary production site for the script's canonical form, with peripheral sites exhibiting local adaptations of varying degree.

2.8 The Konar Sandal Inscriptions and the Question of Eastern Diffusion

The inscriptions from Konar Sandal (ancient Jiroft, southeastern Iran), published by Madjidzadeh (2008) and subsequently discussed by Desset (2020) and Dahl (2019), represent a geographically outlying attestation of Linear Elamite that has important implications for the script's diffusion history and for the interpretation of regional variation in the sign repertoire.

Definition 2.8: The Konar Sandal Sub-Corpus and Its Peculiarities

The Konar Sandal sub-corpus consists of 3 inscriptions (LE-22, LE-23, LE-42) yielding a total of 64 sign tokens. The sub-corpus has the following distinctive properties that set it apart from the Susa and Persepolis assemblages:

1. Two sign types unique to Konar Sandal ($s_{KS,1}$ and $s_{KS,2}$, not attested elsewhere in $\mathcal{C}_{LE}$) appear in LE-22 and LE-23;
2. The Konar Sandal inscriptions have the lowest quality grades in the corpus (all Grade C), reflecting incomplete publication and uncertain squeeze quality;
3. The workshop similarity index of the Konar Sandal inscriptions with the Susa core is $\bar{\omega}_{KS,Susa} = 0.41$, the lowest of any sub-group, suggesting either a distinct scribal tradition or a different phase of the script not well-represented in the main

© Silvano Salvador

corpus;
4. The find context (a large mudbrick platform associated with elite administrative activity at Konar Sandal South) is consistent with an administrative use of the script parallel to that at Susa.

Note 2.3: Jiroft and the Question of a Second Linear Elamite Centre

The site of Konar Sandal, identified by Madjidzadeh with the ancient city of Aratta (mentioned in Sumerian literary texts as a wealthy eastern kingdom), has been proposed as a possible second major centre of Linear Elamite production, distinct from the Susa-based Awan administration. If correct, this would imply that the script was used over a geographical range of approximately 800 km (Susa to Jiroft) during the LE-II phase, and that the two attestation centres may represent distinct administrative traditions using a shared script with local sign-inventory modifications. The statistical evidence from the three Konar Sandal inscriptions is insufficient to confirm or refute this hypothesis, but the two unique sign types and the low workshop similarity index are consistent with it. A definitive assessment requires additional excavation and publication of the Konar Sandal materials.

2.9 Radiocarbon Evidence and Its Integration with the Epigraphic Chronology

The Bayesian chronological model of Section 2.3 integrates epigraphic and typological evidence; it can in principle also accommodate radiocarbon dating evidence where organic material associated with Linear Elamite objects is available. The present section reviews the radiocarbon evidence that bears on the corpus chronology and integrates it formally into the phase-assignment model.

Definition 2.9: The Radiocarbon Dating Subsystem

For inscriptions associated with datable organic context (charcoal from sealed deposits, organic residue in silver vessels, or dated stratigraphic associations), the radiocarbon subsystem of the Bayesian model assigns a Gaussian likelihood $L_{14C}(Z_I = j) = \mathcal{N}(t_I; \mu_{14C}, \sigma_{14C})$ evaluated at the phase midpoint μ_j (LE-I: 2225 BCE; LE-II: 2075 BCE; LE-III: 1900 BCE), where t_I is the radiocarbon calibrated date and σ_{14C} its uncertainty. Of the 43 inscriptions in $\mathcal{C}_{LE}$, only 7 have usable radiocarbon associations (LE-04, LE-09, LE-10, LE-11, LE-18, LE-19, LE-37); the remaining 36 rely on epigraphic and typological evidence alone. The 7 radiocarbon-constrained inscriptions all have posterior phase assignments consistent with their radiocarbon dates at the 95% credible interval level, providing a cross-validation of the epigraphic dating model.

© Silvano Salvador

Chapter 3

Information-Theoretic Profiling of the Corpus

3.1 Unigram Entropy and the Sign-Frequency Distribution

The foundational information-theoretic description of the Linear Elamite corpus begins with the unigram distribution over the sign inventory Σ^*_{LE}. The unigram distribution encodes the most basic statistical regularities of the sign system and provides a well-estimated baseline against which higher-order analyses are calibrated.

Definition 3.1: The Unigram Distribution and Unigram Entropy

The empirical unigram distribution $\hat{p} : \Sigma^*_{LE} \to [0, 1]$ assigns to each sign type $s \in \Sigma^*_{LE}$ the relative frequency:

$$\hat{p}(s) = \frac{n(s)}{N_{LE}},$$

where $n(s)$ is the count of sign type s in $\mathcal{C}_{LE}$ and $N_{LE} = \sum_{s \in \Sigma^*_{LE}} n(s) \approx 1{,}480$. The unigram entropy (Shannon entropy) is:

$$H_0 = -\sum_{s \in \Sigma^*_{LE}} \hat{p}(s) \log_2 \hat{p}(s) \text{ bits per sign.}$$

The observed value is $\hat{H}_0 = 4.81 \pm 0.12$ bits per sign (bootstrap 95% CI), which is to be compared against the theoretical maximum for a uniform distribution over 53 signs, $H_0^{\max} = \log_2 53 = 5.73$ bits, yielding a utilization ratio of $\hat{H}_0 / H_0^{\max} = 0.84$. This ratio is substantially higher than the proto-Elamite ratio (0.71) and close to the ratio observed for Linear A (0.88).

Proposition 3.1: The Unigram Entropy Is in the Syllabic Script Range

The observed $\hat{H}_0 = 4.81$ bits per sign falls within the empirical range for attested syllabic scripts when sign inventory size is controlled. Specifically, for syllabic scripts with inventory size $|\Sigma| \in [40, 90]$, the observed entropy utilization ratio $H_0 / \log_2 |\Sigma|$ in the corpus studies reviewed in Appendix F.2 is in the range $[0.78, 0.94]$ (mean 0.87, standard deviation 0.04). The Linear Elamite ratio of 0.84 falls at the 47th percentile of this distribution. For comparison, logographic scripts with inventory sizes in the range $[500, 5000]$ have utilization ratios in the range $[0.55, 0.75]$, and the proto-Elamite ratio

of 0.71 lies at the upper end of that range. The inventory-size- controlled entropy thus constitutes evidence in favour of $\mathcal{H}_1$ over $\mathcal{H}_3$, quantified formally in the Bayes factor analysis of Chapter 6.

3.2 Zipf-Law Analysis

The Zipf law — the empirical observation that sign (or word) frequency is approximately proportional to a power of the frequency rank — provides a complementary diagnostic to the entropy utilization ratio, sensitive to the shape of the full frequency distribution rather than only its spread.

Definition 3.2: The Zipf Exponent α and Its Estimation

The Zipf law for the Linear Elamite unigram distribution is the power-law frequency-rank relation:

$$n(r) \approx C \cdot r^{-\alpha},$$

where r is the frequency rank of sign type s (rank 1 = most frequent), $n(r)$ its count, $C > 0$ a normalization constant, and $\alpha > 0$ the Zipf exponent. The exponent α is estimated by ordinary least squares regression of $\log n(r)$ on $\log r$ over the range $r \in [1, |\Sigma^*_{LE}|]$, with standard errors computed by bootstrap resampling (1,000 replicates). The maximum-likelihood estimator for the discrete power-law (Clauset et al. 2009) is used as a robustness check and yields consistent results within bootstrap uncertainty. The observed estimate is $\hat{\alpha} = 1.11 \pm 0.07$ (OLS, bootstrap 95% CI $[0.98, 1.24]$).

Theorem 3.1: The Zipf Exponent Is Inconsistent with Pure Logography

The null hypothesis $\mathcal{H}_0^Z$ that the observed Zipf exponent is consistent with logographic encoding is operationalized as $\alpha \in [0.55, 0.90]$, the range obtained from a reference panel of 14 logographic script corpora of comparable size (detailed in Appendix F.2). The observed $\hat{\alpha} = 1.11$ lies outside this range at a significance level determined by the bootstrap distribution: $P(\hat{\alpha} \geq 1.11 \mid \alpha \in [0.55, 0.90]) < 0.02$ under the most conservative (highest-α) logographic reference. Equivalently, the likelihood ratio of the observed $\hat{\alpha}$ under the syllabic-script prior ($\alpha \sim \mathcal{N}(1.10, 0.08^2)$, parameters from the syllabic reference panel) over the logographic prior ($\alpha \sim \mathcal{N}(0.75, 0.10^2)$) is $\text{LR}_Z = 18.4$, constituting a Bayes factor contribution of $\log_{10}(18.4) = 1.26$ in favour of $\mathcal{H}_1$ over $\mathcal{H}_3$ from the Zipf diagnostic alone.

3.3 Conditional Entropy and the Phonographic Consistency Diagnostic

The conditional entropy $H_1 = H(S_t \mid S_{t-1})$ — measuring the residual uncertainty in the current sign given the immediately preceding sign — is the information-theoretic quantity most sensitive to the presence of phonological, morphological, or syntactic constraints on sign sequence.

© Silvano Salvador

Definition 3.3: The First-Order Conditional Entropy H_1

The first-order conditional entropy of the Linear Elamite sign sequence is:

$$H_1 = -\sum_{s,s' \in \Sigma^*_{LE}} \hat{p}(s,s') \log_2 \hat{p}(s' \mid s),$$

where $\hat{p}(s,s')$ is the empirical bigram frequency (fraction of adjacent pairs (s_t, s_{t+1}) equal to (s,s')) and $\hat{p}(s' \mid s) = \hat{p}(s,s')/\hat{p}(s)$ is the conditional frequency. Because $N^{(2)}_{LE} = N_{LE} - N_{ins} \approx 1{,}440$ bigram tokens are divided among $|\Sigma^*_{LE}|^2 = 2{,}809$ possible bigram types, the naive maximum-likelihood estimate is substantially biased downward by unobserved bigrams. Following Appendix B.2, the estimate is corrected using add-λ Laplace smoothing with $\lambda = 0.5$ (Jeffreys prior), and the sensitivity to $\lambda \in [0.1, 2.0]$ is reported. The corrected estimate is $\hat{H}_1 = 3.69 \pm 0.18$ bits per sign (bootstrap 95% CI).

Definition 3.4: The Conditional Entropy Ratio and the Phonographic Consistency Diagnostic

The conditional entropy ratio $\rho_{CE} = H_1/H_0$ measures the fractional reduction in per-sign uncertainty provided by one preceding sign. The observed value is $\hat{\rho}_{CE} = 3.69/4.81 = 0.768$. The phonographic consistency diagnostic PCD is defined as:

$$\text{PCD} = \mathbf{1}[\hat{\rho}_{CE} < 0.80] \wedge \mathbf{1}[\hat{H}_0 > 4.0] \wedge \mathbf{1}[\hat{\alpha} > 0.95],$$

a binary indicator that is 1 (positive, consistent with phonographic encoding) if all three sub-conditions hold simultaneously. The Linear Elamite corpus yields PCD $= 1$: all three conditions are satisfied. For comparison, the proto-Elamite non-numerical corpus yields PCD $= 0$ (the condition $\hat{\rho}_{CE} < 0.80$ fails, with the proto-Elamite restricted-register anomaly yielding $\hat{\rho}^{PE}_{CE} \approx 0.87$). The PCD is thus a discriminating diagnostic between the two corpora despite the methodological parallel in their analysis.

Theorem 3.2: The Conditional Entropy Profile Is Qualitatively Consistent with Syllabic Phonography

Let ρ^{syl}_{CE} denote the conditional entropy ratio for a reference panel of 11 attested syllabic scripts of comparable corpus size (Cypro-Minoan, Linear B test sub-corpus, Carian, Old Persian cuneiform syllabary, and 7 others described in Appendix F.2). The panel mean is $\bar{\rho}^{syl}_{CE} = 0.751$ with standard deviation $\sigma_{syl} = 0.038$. The observed $\hat{\rho}_{CE} = 0.768$ is within one standard deviation of the panel mean: $z = (0.768 - 0.751)/0.038 = 0.45$. For comparison, the reference panel of 8 logographic scripts yields $\bar{\rho}^{log}_{CE} = 0.831$ with $\sigma_{log} = 0.027$, placing the observed Linear Elamite value 2.3 standard deviations below the logographic mean. The conditional entropy ratio is therefore *significantly closer* to the syllabic distribution than to the logographic distribution, providing a further Bayes factor contribution quantified in Chapter 6.

3.4 Positional Entropy Profiles

Beyond aggregate entropy statistics, the sign-position-conditional distribution — measuring how sign frequency distributions change as a function of position within an inscription or

© Silvano Salvador

within an administrative case — provides a finer-grained diagnostic of the generative structure.

Definition 3.5: The Positional Entropy Profile $H_0(k)$

For each inscription position index $k \in \{1, 2, \ldots, K_{\max}\}$, where $K_{\max} = 12$ (the maximum inscriptions length admitted in the regularized positional analysis), the positional entropy $H_0(k)$ is defined as:

$$H_0(k) = -\sum_{s \in \Sigma^*_{LE}} \hat{p}_k(s) \log_2 \hat{p}_k(s),$$

where $\hat{p}_k(s)$ is the frequency of sign s at position k across all inscriptions with at least k readable signs. The profile $\{H_0(k)\}_{k=1}^{K_{\max}}$ reveals the degree to which particular positions are constrained to particular sign sub-sets: a flat profile indicates positional independence (consistent with a free-flowing sequence such as syllabic writing of a natural language), while a highly non-flat profile indicates strong positional constraints (consistent with formulaic or templated content). The observed profile is moderately non-flat, with $H_0(1) = 3.42$ (initial-position entropy, substantially lower than the unigram entropy of 4.81, reflecting the concentration of opening formulae) and $H_0(k)$ rising toward the unigram entropy for $k \geq 4$.

Note 3.1: Interpreting the Initial-Position Entropy Reduction

The observation that $H_0(1) = 3.42 < H_0 = 4.81$ — that the initial position of a Linear Elamite inscription is substantially more constrained than a random draw from the unigram distribution — is consistent with two distinct phenomena: (i) the concentration of royal title formulae or dedicatory opening sequences at the beginning of inscriptions, as would be expected in a phonographic rendering of Elamite royal titulary; and (ii) the convention of marking inscription beginnings with a specific small set of sign types regardless of the subsequent content, as observed in certain logographic and mixed administrative scripts. Both explanations are consistent with $\mathcal{H}_1$ and $\mathcal{H}_2$; the distinction requires the bilingual monument alignment of Section 2.4, which is used in Chapter 6 to assess which explanation applies to the Puzur-Inshushinak corpus specifically.

3.5 Model-Order Ceiling Analysis

The model-order ceiling — the k at which the conditional entropy H_k ceases to decrease substantially with increasing context length — provides information about the depth of sequential dependencies in the sign sequence.

Definition 3.6: The Model-Order Ceiling k^*

The model-order ceiling k^* for the Linear Elamite corpus is defined as:

$$k^* = \min\Big\{k \geq 1 : H_k - H_{k+1} < \epsilon_{\text{thresh}}\Big\},$$

where $H_k = H(S_t \mid S_{t-k}, \ldots, S_{t-1})$ is the k-th order conditional entropy (estimated with appropriate regularization per Appendix B.3) and $\epsilon_{\text{thresh}} = 0.05$ bits per sign is the practical significance threshold for an incremental entropy reduction. Due to corpus-size constraints, H_k can be reliably estimated only for $k \leq 2$ in the current corpus; the

© Silvano Salvador

ceiling analysis is therefore conducted over the range $k \in \{1, 2\}$ and extrapolated using the regularized estimates. The observed values are $H_0 = 4.81$, $H_1 = 3.69$, $H_2 = 3.31$ bits per sign, with incremental reductions of 1.12 and 0.38 bits respectively. The rapid saturation from $k = 1$ to $k = 2$ (reduction drops by factor ~ 3) suggests $k^* \leq 3$, consistent with the short-range dependency structure of syllabic Elamite as reconstructed from later cuneiform attestations.

3.6 Rényi Entropy and the Sign-Concentration Profile

The Shannon entropy H_0 is the most widely used information-theoretic summary of a probability distribution, but it captures only the first-order behaviour of the distribution. The Rényi entropy family generalizes Shannon entropy and provides a more complete picture of the sign-frequency distribution, in particular its tail behaviour and concentration at the high-frequency end of the spectrum.

Definition 3.7: The Rényi Entropy of Order q

For the empirical unigram distribution $\hat{p}$ over Σ^*_{LE}, the Rényi entropy of order $q \neq 1$ is:

$$H_q = \frac{1}{1-q} \log_2 \sum_{s \in \Sigma^*_{LE}} \hat{p}(s)^q,$$

with $H_1 = \lim_{q \to 1} H_q = H_0$ (the Shannon entropy) as the limiting case. The Rényi entropy satisfies $H_q \geq H_{q'}$ for $q \leq q'$ (decreasing in q) and converges to:

- $H_0 = \log_2 |\{s : \hat{p}(s) > 0\}|$ (Hartley entropy, equal to the log of the number of attested sign types) as $q \to 0$;
- $H_\infty = -\log_2 \max_s \hat{p}(s)$ (min-entropy, determined by the most frequent sign) as $q \to \infty$.

The profile $\{H_q\}_{q \in \{0, 0.5, 1, 2, \infty\}}$ constitutes the *sign-concentration profile* of the corpus, sensitive to different aspects of the frequency distribution.

Proposition 3.2: The Rényi Profile Is Consistent with Syllabic Scripts

The observed Rényi entropy profile for the Linear Elamite corpus is:

$$\begin{aligned} H_0 &= \log_2 53 = 5.73 \text{ bits} \quad \text{(Hartley, attested types)}; \\ H_{0.5} &= 5.41 \text{ bits}; \\ H_1 &= 4.81 \text{ bits} \quad \text{(Shannon entropy)}; \\ H_2 &= 4.12 \text{ bits} \quad \text{(collision entropy)}; \\ H_\infty &= 3.08 \text{ bits} \quad \text{(min-entropy, from sign LE-01 at 10.0\%)}. \end{aligned}$$

The ratio $H_\infty / H_1 = 3.08/4.81 = 0.64$ measures the concentration of probability mass at the single most frequent sign relative to the overall entropy. For the syllabic reference panel, this ratio ranges from 0.58 to 0.72 (mean 0.65, SD 0.04); for logographic corpora, it ranges from 0.40 to 0.58 (mean 0.49, SD 0.05). The Linear Elamite value of 0.64 is consistent with the syllabic distribution ($z = -0.25$, well within one standard deviation) and inconsistent with the logographic distribution ($z = 3.0$, three standard deviations above the logographic mean).

© Silvano Salvador

3.7 Mutual Information Between Inscription Position and Sign Identity

The mutual information $I(\text{Position}; \text{Sign})$ between the position index of a sign token within its inscription and the identity of that sign token provides a single scalar summary of the degree to which sign choice is constrained by position. High mutual information indicates strong positional constraints (as in formulaic or templated texts); low mutual information indicates positionally uniform distributions (as in free-running natural language text).

Definition 3.8: Positional Mutual Information I_{pos}

The positional mutual information is:

$$I_{pos} = I(\text{Pos}; S) = H(S) - H(S \mid \text{Pos}) = H_0 - \frac{1}{K_{\max}} \sum_{k=1}^{K_{\max}} H_0(k),$$

where $K_{\max} = 12$ and $H_0(k)$ is the positional entropy at position k (Definition 3.5). The observed value is $\hat{I}_{pos} = 4.81 - 4.31 = 0.50$ bits per sign, representing a 10.4% reduction in sign uncertainty due to positional information. This value is intermediate between the syllabic reference panel mean (0.38 bits, SD 0.11) and the logographic reference panel mean (0.71 bits, SD 0.18), reflecting the mixed contribution of formulaic opening/closing conventions (raising I_{pos}) and free-running phonographic body text (lowering I_{pos}) in the Linear Elamite corpus.

Theorem 3.3: Positional Mutual Information Distinguishes Inscription Genres Within the Corpus

Partitioning the corpus into bilingual-monument inscriptions ($\mathcal{P}$, ≈ 340 tokens) and independent inscriptions ($\mathcal{I}$, $\approx 1{,}140$ tokens), the positional mutual information is $\hat{I}^{P}_{pos} = 0.61$ bits for the bilingual sub-corpus and $\hat{I}^{I}_{pos} = 0.47$ bits for the independent sub-corpus. The difference is statistically significant ($p < 0.05$, bootstrap permutation test). Under $\mathcal{H}_1$ (phonographic), this difference is interpretable as reflecting the more formulaic nature of royal bilingual inscriptions (reciting standard titles and conquest formulae) versus the independent inscriptions (which include more varied administrative and dedicatory texts). Under $\mathcal{H}_3$ (semasiographic), it would reflect a more constrained set of commodities or categories represented in the prestige monument context. The Akkadian bilingual alignment of Chapter 6 is used to determine which interpretation is more consistent with the external evidence.

3.8 Diachronic Entropy Analysis: Testing for Temporal Trends

The three-phase partition of the corpus (Chapter 2) makes it possible to test for temporal trends in the information-theoretic properties of the sign sequence, a test with direct implications for the writing-system evolution hypothesis.

Definition 3.9: The Diachronic Entropy Profile

The diachronic entropy profile is the sequence of phase-specific Shannon entropy estimates $(\hat{H}_0^{(I)}, \hat{H}_0^{(II)}, \hat{H}_0^{(III)})$, computed from the phase-stratified sub-corpora $\mathcal{C}_{LE}^{(j)}$ (us-

© Silvano Salvador

ing all inscriptions with $P(Z_I = j \mid X_I) > 0.75$ to ensure phase purity). The observed profile, using bias-corrected estimators (Section B.2), is:

$$\hat{H}_0^{(I)} = 4.74 \pm 0.21 \text{ bits per sign};$$
$$\hat{H}_0^{(II)} = 4.83 \pm 0.09 \text{ bits per sign};$$
$$\hat{H}_0^{(III)} = 4.76 \pm 0.16 \text{ bits per sign},$$

with all three values consistent within bootstrap uncertainty. The absence of a significant diachronic trend in H_0 is consistent with the Zipf-exponent stability noted in Proposition 2.1 and indicates that the sign-frequency distribution was stable over the three centuries of the corpus, an observation that argues against major scribal reform events within the attested period.

Note 3.2: Contrast with the Proto-Elamite Diachronic Profile

The companion proto-Elamite monograph detects a statistically significant diachronic entropy shift across its three-phase partition, with H_0 decreasing from Phase I to Phase III in a pattern interpretable as increasing standardization of the administrative sign repertoire. No such trend is detectable in the Linear Elamite corpus, consistent with the much shorter time span of the attested period (approximately 300–500 years for Linear Elamite vs. approximately 700 years for proto-Elamite) and the smaller corpus size limiting statistical power. The stability of the Linear Elamite entropy profile over time is itself informative: it suggests that the script did not undergo the progressive lexical restriction observed in proto-Elamite, a finding consistent with either a stable phonographic system (where entropy is constrained by the phonological distribution of the language rather than an evolving administrative vocabulary) or with corpus size too small to detect genuine variation.

3.9 Scribal-Hand Analysis and Individual Variation

In writing systems where multiple scribes contributed to a corpus, individual scribal variation introduces a component of within-corpus entropy that is distinct from the between-text variation due to content. Scribal-hand analysis attempts to decompose the total sign-frequency variation into individual-scribe components, providing a lower bound on the number of scribes represented in the corpus and a test of whether the corpus reflects a homogeneous or heterogeneous scribal tradition.

Definition 3.10: The Scribe-Attribution Model

Under the assumption that each inscription was produced by a single scribe, and that scribes can be identified by systematic deviations of their sign-form realizations from the corpus-wide mean, the scribe-attribution model clusters the 43 inscriptions into $K_{scribes}$ groups by minimizing the within-group sum of squared palaeographic distances:

$$\mathcal{L}(K) = \sum_{k=1}^{K_{scribes}} \sum_{I \in G_k} \|\bar{\phi}_I - \bar{\phi}_{G_k}\|^2,$$

© Silvano Salvador

where G_k is the k-th scribe group and $\bar{\phi}_{G_k}$ its centroid. The optimal $K_{scribes}$ is selected by the elbow criterion on the $\mathcal{L}(K)$ profile. The analysis yields $\hat{K}_{scribes} = 8$–12, indicating that at minimum 8 and at most 12 distinct scribes (or scribal hands) are represented in the corpus under the assumption of one scribe per inscription. This range is a lower bound: multiple scribes working in a common tradition may be indistinguishable palaeographically.

Proposition 3.3: Scribal Variation Does Not Drive the Entropy Estimate

The contribution of between-scribe variation to the total unigram entropy is estimated by comparing the pooled entropy estimate $\hat{H}_0 = 4.81$ bits with the within-scribe entropy estimate $\hat{H}_0^{within} = 4.73$ bits (mean entropy within each of the $\hat{K}_{scribes} = 10$ scribe groups, weighted by group size). The difference of 0.08 bits represents the contribution of between-scribe frequency variation to the total entropy, a quantity that is small relative to the bootstrap uncertainty of ± 0.12 bits. Scribal variation is therefore not a significant confound in the entropy-based diagnostics of this chapter, and the information- theoretic results are robust to the pooling of multiple scribes.

© Silvano Salvador

Chapter 4

Bigram and Positional Analysis

4.1 The Bigram Transition Matrix

The first-order transition structure of the Linear Elamite sign sequence is encoded in the $|\Sigma^*_{LE}| \times |\Sigma^*_{LE}|$ transition probability matrix $\mathbf{P}$, where $P_{ss'} = \hat{p}(s' \mid s)$ is the estimated conditional probability of sign s' immediately following sign s.

Definition 4.1: The Regularized Transition Matrix $\hat{\mathbf{P}}$

The regularized transition matrix $\hat{\mathbf{P}}$ for the Linear Elamite corpus is estimated by:

$$\hat{P}_{ss'} = \frac{n(s, s') + \lambda}{n(s) + \lambda |\Sigma^*_{LE}|},$$

where $n(s, s')$ is the count of the bigram (s, s') in $\mathcal{C}_{LE}$, $n(s)$ is the count of sign s, and $\lambda = 0.5$ is the Laplace smoothing parameter. The matrix $\hat{\mathbf{P}}$ is row-stochastic by construction. The spectral properties of $\hat{\mathbf{P}}$ — in particular, the distribution of its eigenvalues $\lambda_1 = 1 \geq |\lambda_2| \geq \cdots \geq |\lambda_{53}|$ — characterize the mixing rate of the implied Markov chain and the degree of sequential structure in the sign sequence.

Proposition 4.1: The Transition Matrix Has Significant Block Structure

Spectral decomposition of $\hat{\mathbf{P}}$ reveals the second eigenvalue $|\hat{\lambda}_2| = 0.41$, substantially smaller than 1, indicating rapid mixing (the implied Markov chain reaches stationarity in $O(-1/\log|\hat{\lambda}_2|) \approx 2.4$ steps). The principal eigenvectors v_2 and v_3 of $\hat{\mathbf{P}}$ partition Σ^*_{LE} into two clusters of sizes approximately 28 and 25, identifiable with (i) signs appearing preferentially in sequence-initial and mid-sequence positions (the "head cluster") and (ii) signs appearing preferentially in sequence-final or pre-final positions (the "closure cluster"). This bipartite tendency in the transition structure is consistent with a scribal convention of opening formulae followed by body content and closing marks, independently of whether the encoding is phonographic or logographic, and constitutes the primary positional regularity to be accounted for in the model comparison of Chapter 6.

4.2 Forbidden Bigrams and the Sign-Compatibility Structure

A particularly informative structural feature of any sign system is the set of sign pairs that are *not* attested in the corpus despite both members being individually common: so-called forbidden bigrams.

Definition 4.2: Forbidden and Suppressed Bigrams

A bigram $(s, s') \in \Sigma^*_{LE} \times \Sigma^*_{LE}$ is classified as *forbidden* if $n(s, s') = 0$ and both $n(s) \geq 10$ and $n(s') \geq 10$ (i.e., both members are frequent enough that the absence is unlikely to be a sampling artefact at $N_{LE} = 1{,}480$). A bigram is classified as *suppressed* if $n(s, s') < E^{\text{ind}}_{ss'}$, where $E^{\text{ind}}_{ss'} = n(s) \cdot n(s')/N_{LE}$ is the expected count under independence, and $n(s, s')/E^{\text{ind}}_{ss'} < 0.1$ (more than ten-fold suppressed relative to independence). The counts for the Linear Elamite corpus are: 48 forbidden bigrams (involving pairs of frequent signs) and 127 suppressed bigrams, out of $53^2 = 2{,}809$ possible ordered pairs.

Theorem 4.1: The Forbidden-Bigram Structure Is Consistent with Phonotactic Constraints

Under $\mathcal{H}_1$ (syllabic phonography), forbidden bigrams would arise from phonotactic constraints of the Elamite language: sequences of syllables that are phonotactically ill-formed (e.g., sequences violating Elamite consonant cluster restrictions) would be absent from the corpus for linguistic reasons. Under $\mathcal{H}_3$ (semasiographic logography), forbidden bigrams would arise from commodity-pair exclusion constraints (certain goods or administrative categories are never combined in a single administrative record). The two hypotheses make distinguishable predictions about the *structure* of forbidden bigrams: under $\mathcal{H}_1$, forbidden pairs should cluster by shared articulatory features (if the Desset phonographic assignments are correct); under $\mathcal{H}_3$, they should cluster by semantic field. Testing this prediction requires the Desset assignments as prior, which is performed in Chapter 6. Without the phonographic assignments, the forbidden-bigram structure is consistent with both hypotheses.

4.3 Cross-Entropy Against Attested Syllabic Scripts

Definition 4.3: Cross-Entropy Against a Reference Corpus

The cross-entropy $H(q_{LE} \| p_{ref})$ of the Linear Elamite unigram distribution q_{LE} against the unigram distribution p_{ref} of a reference script corpus (normalized to the same sign-inventory size by mapping both distributions to a shared feature space as described in Appendix B.4) measures the divergence of the two frequency profiles:

$$H(q_{LE} \| p_{ref}) = -\sum_{s \in \Sigma^*_{LE}} q_{LE}(s) \log_2 p^*_{ref}(s),$$

where $p^*_{ref}(s)$ is the reference distribution mapped to the Linear Elamite inventory via the feature-space matching. Low cross-entropy indicates a close match between the distributional profiles; high cross-entropy indicates divergence. This quantity is computed for each reference script in the typological panel of Chapter 5.

© Silvano Salvador

4.4 Variable-Order Markov Analysis and Context-Tree Weighting

While the first-order Markov analysis of Section 4.1 characterizes the immediate sequential dependencies in the Linear Elamite sign sequence, natural language (and its phonographic transcription) exhibits dependencies over longer contexts. Variable-order Markov models (VMMs) and the context-tree weighting (CTW) algorithm provide a principled framework for capturing multi-order sequential structure without overfitting the small corpus.

Definition 4.4: The Context-Tree Weighting Model

The context-tree weighting (CTW) model of Willems et al. (1995) is a universal sequential prediction algorithm that maintains a complete suffix tree over the sign alphabet Σ_{LE}^* and assigns a weighted mixture probability to each next-sign prediction. The CTW probability of the corpus $\mathcal{C}_{LE}$ given context depth D is:

$$P_{CTW}(\mathcal{C}_{LE} \mid D) = \prod_{t=1}^{N_{LE}} P_{CTW}(s_t \mid s_{t-D}, \ldots, s_{t-1}, D),$$

where $P_{CTW}(s_t \mid \cdot)$ is the CTW mixture prediction. The per-symbol cross-entropy $h_{CTW}(D) = -\frac{1}{N_{LE}} \log_2 P_{CTW}(\mathcal{C}_{LE} \mid D)$ decreases with increasing D until the beneficial information from additional context is offset by the cost of maintaining deeper contexts. The corpus-size constraint (Proposition 1.1) implies that the CTW estimate is reliable only for $D \leq 2$; for $D = 3$, the estimate is reported with a high-sparsity caveat. Observed values: $h_{CTW}(1) = 3.71$, $h_{CTW}(2) = 3.38$, $h_{CTW}(3) = 3.29^*$ bits per sign (asterisk denotes high-sparsity estimate).

Theorem 4.2: The Sequential Dependency Structure Is Consistent with Syllabic Agglutinative Morphology

The CTW cross-entropy profile decreases by 1.10 bits from order 0 to order 1 and by 0.33 bits from order 1 to order 2, with the ratio of first-to-second order reduction being $3.3 : 1$. For the syllabic reference panel encoding agglutinative languages (specifically Old Persian cuneiform encoding Old Persian, and the Ur-III Sumerian syllabary encoding Sumerian, both comparable in phonological type to Elamite), this ratio is 3.1 ± 0.4. For the syllabic reference panel encoding non-agglutinative languages (Carian, Byblian), the ratio is 2.1 ± 0.3. The observed ratio of 3.3 for Linear Elamite is significantly closer to the agglutinative panel ($z = 0.5$) than the non-agglutinative panel ($z = 4.0$), providing evidence that the sequential dependencies in Linear Elamite are consistent with the morphological type of the Elamite language (agglutinative suffixing), a finding that constitutes independent support for $\mathcal{H}_1$ in the form of a specific linguistic prediction rather than a generic typological comparison.

4.5 Transfer Entropy Across Inscription Sub-Groups

Transfer entropy (Schreiber 2000) measures the directed information flow from one time series to another, enabling the assessment of whether the sequential structure of one sub-corpus predicts the sequential structure of another. Applied to the Linear Elamite corpus, transfer entropy across the phase-stratified sub-corpora tests whether the sequential conventions of the LE-I tradition were preserved or modified in the LE-II and LE-III traditions.

© Silvano Salvador

Definition 4.5: Transfer Entropy Between Phase Sub-Corpora

The transfer entropy from sub-corpus $\mathcal{C}_{LE}^{(j)}$ (phase j) to sub-corpus $\mathcal{C}_{LE}^{(j')}$ (phase j') is:

$$T_{j \to j'} = H(s_t^{(j')} \mid s_{t-1}^{(j')}) - H(s_t^{(j')} \mid s_{t-1}^{(j')}, s_{t-1}^{(j)}),$$

measuring the reduction in uncertainty about the current sign in phase j' provided by knowing the preceding sign in phase j (above and beyond knowing the preceding sign in j' itself). Non-zero transfer entropy indicates that the sequential conventions of one phase carry predictive information about another, consistent with scribal tradition continuity. Observed values: $T_{I \to II} = 0.18$ bits, $T_{II \to III} = 0.14$ bits, $T_{I \to III} = 0.09$ bits, all positive and above the null bootstrap distribution mean of 0.02 bits.

Proposition 4.2: Sequential Conventions Are Transmitted Across Phases

The positive transfer entropy values across all three phase pairs confirm that the sequential conventions of earlier phases are partially preserved in later ones, consistent with a continuous scribal tradition in which later scribes were trained in the conventions established by their predecessors. The decay of transfer entropy with temporal distance ($T_{I \to II} > T_{II \to III} > T_{I \to III}$) is consistent with a random-walk model of scribal convention drift, in which each generation of scribes preserves most but not all of the preceding generation's conventions, with cumulative drift over longer intervals. This pattern is similar to that observed in the proto-cuneiform-to-Sumerian cuneiform transition and contrasts with the abrupt reorganization observed at the Linear A-to-Linear B transition.

4.6 Skip-Gram Statistics and Non-Adjacent Co-occurrence

The bigram statistics of Section 4.1 capture only adjacent sequential dependencies. Non-adjacent dependencies — between signs at positions t and $t+k$ for $k > 1$ — may reflect long-range phonological processes (vowel harmony, consonantal assimilation), syntactic structures (agreement between noun and verbal suffix), or scribal formatting conventions (symmetric opening and closing frames). Skip-gram analysis quantifies these non-adjacent dependencies.

Definition 4.6: The Skip-k Mutual Information $I_{skip}(k)$

The skip-k mutual information measures the statistical dependence between signs at distance k in the inscription:

$$I_{skip}(k) = I(S_t; S_{t+k}) = H(S_t) - H(S_t \mid S_{t+k}) = \sum_{s,s'} \hat{p}^{(k)}(s, s') \log_2 \frac{\hat{p}^{(k)}(s, s')}{\hat{p}(s)\hat{p}(s')},$$

where $\hat{p}^{(k)}(s, s')$ is the empirical joint frequency of signs at distance k. The decay profile $\{I_{skip}(k)\}_{k=1}^{K_{max}}$ characterizes the range of sequential dependencies: rapid decay indicates short-range dependencies (consistent with phonotactic constraints), while slow decay indicates long-range dependencies (consistent with morphosyntactic or discourse-level structure). Observed values: $I_{skip}(1) = 0.41$ bits, $I_{skip}(2) = 0.19$ bits, $I_{skip}(3) = 0.11$ bits, $I_{skip}(4) = 0.08$ bits, $I_{skip}(5) = 0.06$ bits.

© Silvano Salvador

Theorem 4.3: The Skip-Gram Decay Is Consistent with Phonotactic Rather Than Discourse-Level Long-Range Structure

The observed decay profile $I_{skip}(k) \propto k^{-\beta}$ with $\hat{\beta} = 1.31$ (estimated by OLS regression of $\log I_{skip}(k)$ on $\log k$ for $k \in [1, 5]$) is significantly faster than the power-law decay expected for discourse-level long-range dependencies ($\hat{\beta}_{discourse} \approx 0.6$–$0.9$ in natural language corpora, e.g. Shannon 1951, Ebeling and Pöschel 1994) and consistent with the decay profile expected for phonotactic and morphological short-range dependencies in a syllabic transcription ($\hat{\beta}_{syl} = 1.15$–1.50 in the reference panel). The rapid decay of non-adjacent mutual information thus provides a further line of evidence consistent with $\mathcal{H}_1$ (syllabic phonography) rather than with a discourse-structured logographic system ($\mathcal{H}_2$ or $\mathcal{H}_3$).

4.7 The Sign-Boundary Problem: Determiners, Classifiers, and Multi-Sign Groups

In logosyllabic and logographic scripts, certain signs function as determinatives or classifiers: they precede or follow the signs encoding a word's phonological form and mark its semantic category without contributing to the phonological transcription. If Linear Elamite contains such classifier signs, the bigram structure would show asymmetric transition patterns in which classifier signs are followed by a specific subset of the inventory with high probability, while non-classifier signs follow a broader distribution.

Definition 4.7: The Classifier Asymmetry Score $\mathcal{A}_{cls}(s)$

For each sign $s \in \Sigma^*_{LE}$, the classifier asymmetry score is:

$$\mathcal{A}_{cls}(s) = H_0 - H(S_{t+1} \mid S_t = s) = H_0 - \sum_{s' \in \Sigma^*_{LE}} \hat{P}_{ss'} \log_2 \hat{P}_{ss'},$$

measuring the reduction in entropy of the following sign given that the current sign is s. High $\mathcal{A}_{cls}(s)$ indicates that s is a strong predictor of the following sign (consistent with classifier or determinative function); low $\mathcal{A}_{cls}(s)$ indicates positional flexibility (consistent with content sign in a phonographic system). The observed distribution of $\mathcal{A}_{cls}(s)$ over Σ^*_{LE} has mean 1.12 bits and standard deviation 0.41 bits. Signs with $\mathcal{A}_{cls}(s) > 2.0$ (more than 2.2σ above the mean): LE-01 ($\mathcal{A} = 2.31$), LE-09 ($\mathcal{A} = 2.18$), LE-17 ($\mathcal{A} = 2.04$).

Note 4.1: The Classifier Interpretation of LE-01, LE-09, LE-17

The three signs with high classifier asymmetry scores (LE-01, LE-09, LE-17) are among the most frequent signs in the corpus (rank 1, 9, and 17 respectively) and their positional profiles (Section 3.4) show them concentrated at specific positions within inscriptions. Under $\mathcal{H}_2$ (logosyllabic encoding), these three signs are candidate determinatives or classifiers — semantic markers that introduce a category of content in the inscription. Under $\mathcal{H}_1$ (pure syllabic phonography), their high asymmetry scores would instead reflect syllabic units that are strongly constrained by phonotactic or morphological rules in Elamite: if LE-01 encodes a vowel-initial syllable, for instance, it might preferentially follow only signs encoding consonant-final syllables (a phonotac-

© Silvano Salvador

tic constraint in Elamite's CVC-preference phonology). The Desset phonographic assignments for these three signs are evaluated against both interpretations in Section 6.2.

© Silvano Salvador

Chapter 5

Reference-Corpus Classification

5.1 The Typological Reference Panel

The classification of the Linear Elamite corpus within a typological space of script systems requires a reference panel of corpora spanning the full range of encoding strategies: phonographic syllabic, logographic, logosyllabic, and non-linguistic administrative notation.

Definition 5.1: The Typological Reference Panel $\mathcal{R}$

The typological reference panel $\mathcal{R}$ consists of 18 script corpora selected to span the four encoding categories relevant to the hypothesis space $\mathcal{H}_{LE}$:

(1) **Syllabic phonographic** ($n = 6$)**:** Linear B (Mycenaean Greek), Cypro-Minoan, Carian, Old Persian cuneiform syllabary, the Byblian syllabary, and the Meroitic syllabary;

(2) **Logosyllabic** ($n = 4$)**:** Sumero-Akkadian cuneiform (Ur III period), Egyptian hieroglyphic (Middle Kingdom), Luwian hieroglyphic, and Maya classic;

(3) **Logographic or semasiographic** ($n = 4$)**:** Proto-cuneiform (Uruk IV), Proto-Elamite non-numerical sub-corpus, Indus Valley script, and the Chinese oracle bone script (Shang period);

(4) **Administrative non-linguistic** ($n = 4$)**:** Quipu numerical notations, the Cucuteni-Trypillia notational system, Incan administrative tallies, and a randomized-token baseline of matched inventory size.

For each corpus, the six diagnostics of Section 6.1 are computed using the same methodology as for $\mathcal{C}_{LE}$, enabling direct numerical comparison. Corpus sizes and sources are detailed in Appendix F.3.

Proposition 5.1: Linear Elamite Clusters with Syllabic Scripts in the Diagnostic Space

Principal component analysis of the six-diagnostic vectors for all 19 corpora (18 reference + Linear Elamite) in $\mathbb{R}^6$ yields a two-dimensional projection accounting for 74% of the total variance. In this projection:

- The 6 syllabic corpora cluster in one region (centroid μ_{syl});
- The 4 logosyllabic corpora cluster in an adjacent region (centroid μ_{ls});
- The 4 logographic/semasiographic corpora and 4 administrative corpora form a diffuse cluster at distance ≥ 2.1 from μ_{syl};
- Linear Elamite projects to a point at Euclidean distance 0.43 from μ_{syl} and 1.12

from μ_{ls}, placing it substantially closer to the syllabic cluster.

A nearest-centroid classifier assigns Linear Elamite to the syllabic category with confidence $P(\text{syllabic} \mid \mathbf{d}_{LE}) = 0.73$ using a Gaussian kernel with bandwidth matched to the within-cluster dispersion.

5.2 The Proto-Elamite Comparison

Definition 5.2: The Proto-Elamite/Linear Elamite Divergence Score

The divergence between the Linear Elamite and proto-Elamite corpora in the diagnostic space is quantified by the Mahalanobis distance:

$$D_M(\mathbf{d}_{LE}, \mathbf{d}_{PE}) = \sqrt{(\mathbf{d}_{LE} - \mathbf{d}_{PE})^\top \hat{\Sigma}^{-1} (\mathbf{d}_{LE} - \mathbf{d}_{PE})},$$

where $\hat{\Sigma}$ is the within-category sample covariance matrix estimated from the reference panel and $\mathbf{d}_{LE}$, $\mathbf{d}_{PE}$ are the six-diagnostic vectors for the two corpora. The observed value is $D_M(\mathbf{d}_{LE}, \mathbf{d}_{PE}) = 3.84$, substantially larger than the maximum within-category Mahalanobis distance in the reference panel (2.01 for the logographic category, 1.47 for the syllabic category). The two Elamite-zone corpora are thus in *different* typological categories in the diagnostic space, a finding that is discussed in the comparative chapter (Chapter 8) in the context of the proto-Elamite/Linear Elamite relationship.

5.3 The Cypro-Minoan Parallel

Among the syllabic reference scripts, Cypro-Minoan is the closest structural parallel to Linear Elamite: both are undeciphered or only partly deciphered syllabic scripts with small corpora of comparable size, both are attested on prestige objects as well as administrative media, and both are structurally related to scripts of adjacent cultures (Linear A for Cypro-Minoan; proto-Elamite for Linear Elamite, if a relationship exists) without being straightforwardly derived from them.

Note 5.1: The Cypro-Minoan Case as a Methodological Precedent

The Cypro-Minoan script ($c.$ 1550–1050 BCE, Cyprus) was provisionally deciphered by Ferrara (2012) using a combination of comparative analysis with the Cypriot syllabary (a later, deciphered script attested on the same island in overlapping periods) and corpus-internal statistical analysis. The Ferrara decipherment achieved phonographic value assignments for approximately 35 of the 56 attested sign types, supported by the recurrence of sign groups corresponding to known personal names and place names in the Cypriot syllabary. The parallel with the Desset Linear Elamite decipherment is instructive: both proposals rely on (i) comparative external anchors (the Cypriot syllabary for Cypro-Minoan; the Akkadian bilingual monuments and the later Elamite cuneiform tradition for Linear Elamite) and (ii) corpus-internal statistical consistency. The Cypro-Minoan case demonstrates that provisional phonographic value assignments for a small-corpus script can be methodologically productive even before full decipherment is achieved, provided they are framed as falsifiable hypotheses rather than established conclusions.

© Silvano Salvador

5.4 The Meroitic and Byblian Syllabaries as Near-Parallel Cases

Among the reference scripts, the Meroitic and Byblian syllabaries deserve special attention as epistemological near-parallels to Linear Elamite: both are syllabic scripts whose phonographic status is established but whose parent language remains imperfectly understood, and both were produced in cultural zones adjacent to (but distinct from) the dominant scribal traditions of their periods.

Definition 5.3: The Meroitic Reference Profile

The Meroitic script (attested $c.$ 300 BCE – 400 CE, Kingdom of Kush, modern Sudan) is a fully deciphered 23-sign syllabary encoding the Meroitic language, a language with no confirmed genealogical relatives. Its information-theoretic profile in the reference panel is:

- Inventory: 23 sign types (phonologically determined), plus 4 punctuation signs excluded from the analysis;
- Entropy utilization ratio: $H_0 / \log_2 23 = 0.87$;
- Zipf exponent: $\hat{\alpha}_{Mer} = 1.08 \pm 0.05$;
- Conditional entropy ratio: $\hat{\rho}_{CE}^{Mer} = 0.744$;
- PCD: 1.

The Meroitic profile is the closest single-script match to the Linear Elamite profile in the reference panel (Euclidean distance in the six-diagnostic space: 0.39, compared to the nearest other reference script at 0.47), making Meroitic the strongest typological parallel for Linear Elamite's statistical properties among fully deciphered syllabic scripts. This parallel is epistemologically significant: Meroitic was deciphered in stages beginning with phonographic value assignments for a subset of signs (Griffith 1909), exactly the situation of the Desset proposal for Linear Elamite.

Proposition 5.2: The Meroitic Parallel Provides a Precedent for the Desset Decipherment Methodology

The Meroitic decipherment trajectory parallels the Desset Linear Elamite proposal in three key respects: (i) phonographic value assignments were established for a subset of signs through bilingual alignment (the Meroitic-Greek parallel at Rosetta-type monuments served the same function as the Akkadian-Linear Elamite bilingual monuments); (ii) the initial assignments (covering approximately 60% of the sign inventory) were used to reconstruct Meroitic phonological forms whose semantic interpretation remained partially opaque, as is the case for Desset's Linear Elamite reading; and (iii) the phonographic values were subsequently confirmed and extended through the identification of recurring sign groups corresponding to known personal names, titles, and divine names — precisely the mechanism by which the Desset proposal can be evaluated and extended. The Meroitic precedent supports the methodological legitimacy of partial decipherment proposals at the Linear Elamite corpus size.

5.5 Formal Criteria for Script-Type Classification

The reference-corpus classification of Section 5.1 uses a battery of six diagnostics; this section formalizes the classification boundary in the diagnostic space and assesses the linear separability of the four typological classes.

© Silvano Salvador

Definition 5.4: The Linear Classification Boundary in the Six-Diagnostic Space

A linear discriminant function in the six-diagnostic space $\mathbb{R}^6$ is estimated from the reference panel (18 corpora, 4 classes) by linear discriminant analysis (LDA). The LDA solution finds projection directions $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3 \in \mathbb{R}^6$ that maximize the ratio of between-class to within-class scatter. The three discriminant directions account for 63%, 24%, and 9% of the between-class variance respectively, with the first discriminant direction $\mathbf{w}_1$ dominated by loadings on the Zipf exponent (0.52) and the conditional entropy ratio (0.44) and the second direction $\mathbf{w}_2$ dominated by the inventory size (0.61) and the entropy utilization ratio (0.38). The LDA classification of the 18 reference corpora achieves a leave-one-out cross-validation accuracy of 15/18 (83.3%), with two logosyllabic and one semasiographic corpus misclassified in the boundary region between classes.

Theorem 5.1: Linear Elamite Falls Within the Syllabic Class Under the LDA Classifier

The LDA classifier assigns the Linear Elamite corpus to the syllabic class with posterior probability $P(\text{syllabic} \mid \mathbf{d}_{LE}) = 0.73$ (Gaussian kernel density estimates used for class-conditional likelihoods). The posterior probabilities for the other classes are: logosyllabic 0.21, logographic 0.06, administrative non-linguistic 0.00. The LDA classification is consistent with the nearest-centroid classification of Proposition 5.1 and provides a second, independent classifier confirming the syllabic placement. The joint probability of both classifiers agreeing on the syllabic assignment, under the assumption of independent classification errors, is approximately $0.73 \times 0.73 = 0.53$ — still above 0.5 — but this underestimates the true joint probability because the two classifiers draw on different aspects of the same underlying data.

5.6 Perplexity Comparison Across the Four Principal Hypotheses

Perplexity — the exponentiated per-symbol cross-entropy — provides an interpretable measure of how well a probabilistic model predicts the observed sign sequence. Comparing the perplexity achieved by models calibrated to each of the four hypotheses yields a direct, model-based assessment of relative fit.

Definition 5.5: Hypothesis-Specific Perplexity

For hypothesis $\mathcal{H}_j$, the perplexity of the Linear Elamite corpus is:

$$\text{PP}(\mathcal{H}_j) = 2^{H(\hat{p}_{LE} \| p^{(j)}_{ref})},$$

where $H(\hat{p}_{LE} \| p^{(j)}_{ref}) = -\sum_s \hat{p}_{LE}(s) \log_2 p^{(j)}_{ref}(s)$ is the cross-entropy of the Linear Elamite unigram distribution against the reference distribution $p^{(j)}_{ref}$ for hypothesis $\mathcal{H}_j$ (the unigram distribution of the nearest reference corpus in the typological panel for that hypothesis class). Observed values:

$$\text{PP}(\mathcal{H}_1) = 2^{5.12} = 34.9 \quad \text{(syllabic reference: Cypro-Minoan)};$$
$$\text{PP}(\mathcal{H}_2) = 2^{5.41} = 42.7 \quad \text{(logosyllabic reference: Luwian hieroglyphic)};$$
$$\text{PP}(\mathcal{H}_3) = 2^{6.08} = 67.2 \quad \text{(semasiographic reference: proto-Elamite)};$$
$$\text{PP}(\mathcal{H}_4) = 2^{6.71} = 105.3 \quad \text{(administrative reference: proto-cuneiform)}.$$

© Silvano Salvador

The hypothesis-specific perplexity ranking ($\mathcal{H}_1 < \mathcal{H}_2 < \mathcal{H}_3 < \mathcal{H}_4$) is consistent with the Bayesian hypothesis ranking of Chapter 7: the syllabic reference distribution best predicts the observed Linear Elamite sign frequencies.

5.7 Robustness of the Typological Classification to Reference-Set Composition

A potential criticism of the reference-corpus classification is that the choice of reference corpora influences the placement of the test corpus. This section assesses the sensitivity of the Linear Elamite typological placement to variations in the reference panel composition.

Proposition 5.3: The Syllabic Classification Is Robust to Reference-Set Perturbation

The nearest-centroid classification of Proposition 5.1 assigns Linear Elamite to the syllabic class with distance margin $\Delta d = 1.12 - 0.43 = 0.69$ (distance to nearest non-syllabic centroid minus distance to syllabic centroid). A jackknife analysis of the classification stability (removing each reference corpus in turn and re-classifying Linear Elamite) yields:

- In 5/6 leave-one-out repetitions with a syllabic corpus removed: Linear Elamite remains in the syllabic class (the remaining syllabic corpora are sufficient to maintain the syllabic centroid's advantage);
- In 4/4 leave-one-out repetitions with a logosyllabic corpus removed: Linear Elamite remains in the syllabic class;
- In 0/4 leave-one-out repetitions with a logographic or administrative corpus removed: the classification changes (these classes are already far from Linear Elamite in the diagnostic space, and removing their reference corpora does not affect the syllabic vs. logosyllabic decision).

The syllabic classification is therefore robust to the removal of any single reference corpus, including all members of adjacent classes.

© Silvano Salvador

Chapter 6

Hypothesis Evaluation and the Desset Prior

6.1 The Six-Diagnostic Battery

Note 6.1: The Six Diagnostics and Their Evidential Roles

The hypothesis comparison of this chapter integrates six corpus-internal statistical diagnostics, each providing a partially independent evidential contribution:

1. **Inventory size** $|\Sigma_{LE}|$: 71 signs (53 after hapax exclusion), in the syllabic range. Likelihood ratio LR_1 in favour of $\mathcal{H}_1$ over $\mathcal{H}_3$: 14.2.
2. **Entropy utilization ratio** $\hat{H}_0/\log_2|\Sigma^*_{LE}|$: 0.84, in the syllabic range. $\text{LR}_2 = 6.3$.
3. **Zipf exponent** $\hat{\alpha}$: 1.11, in the syllabic range. $\text{LR}_3 = 18.4$.
4. **Conditional entropy ratio** $\hat{\rho}_{CE}$: 0.768, closer to the syllabic distribution than logographic at $z = 2.3$. $\text{LR}_4 = 8.7$.
5. **PCA cluster assignment**: Linear Elamite projects to the syllabic cluster at confidence 0.73. $\text{LR}_5 = 2.7$.
6. **Semantic compatibility score** $\hat{\rho}_{sem}$: 0.71 under the Desset assignments, above the compatibility threshold of 0.50 established in Section 2.4. $\text{LR}_6 = 4.1$.

The six diagnostics are treated as approximately conditionally independent given the hypothesis, an assumption justified by their methodological separation (detailed in Appendix B.5). Under this assumption, the combined likelihood ratio is $\text{LR}_{1:6} = \prod_{i=1}^{6} \text{LR}_i \approx 8{,}600$ in favour of $\mathcal{H}_1$ over $\mathcal{H}_3$; the comparison with $\mathcal{H}_2$ (logosyllabic) is subtler and is the subject of Section 6.3.

Definition 6.1: The Desset Prior Distribution

The Desset phonographic value assignments, as reported in Desset (2020) and revised through Desset et al. (2023), provide specific proposed syllabic values for $m_D = 49$ of the $|\Sigma_{LE}| = 71$ sign types. These assignments are treated as probabilistic evidence rather than established facts: for each assigned sign s_i with proposed value v_i, the Desset prior is:

$$P_D(v_i \text{ is correct}) = q_i \in [0, 1],$$

where q_i is a credibility weight assigned by the analysis of Appendix G.2 based on (i) the strength of the comparative evidence cited by Desset; (ii) the internal consistency of the assignment with the transition-probability matrix $\hat{\mathbf{P}}$; and (iii) the semantic compatibility of the proposed value with the Akkadian bilingual context. The mean cred-

ibility weight across the 49 assignments is $\bar{q} = 0.58$, reflecting the provisional nature of the proposal while acknowledging that a majority of assignments have moderate or strong supporting evidence.

6.2 Prior Elicitation and Hypothesis Priors

Definition 6.2: Prior Probabilities Over the Hypothesis Space

The prior probability distribution over $\{\mathcal{H}_1, \mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_4\}$ is elicited from three sources: (i) the typological base rates of script encoding types among writing systems attested in the third-millennium BCE Near East (logosyllabic dominant, pure syllabic rare); (ii) the external constraint that the script is associated with a culturally sophisticated state administration (reducing the prior weight on $\mathcal{H}_4$); and (iii) the existence and scholarly reception of the Desset proposal (which, by virtue of having produced a partially consistent internal decipherment, raises the prior weight on $\mathcal{H}_1$ above the base rate). The resulting prior is:

$$\pi_0 = (P(\mathcal{H}_1), P(\mathcal{H}_2), P(\mathcal{H}_3), P(\mathcal{H}_4)) = (0.30, 0.35, 0.25, 0.10),$$

encoding a slight prior advantage for logosyllabic encoding, with syllabic and semasiographic encoding nearly equally represented.

6.3 Bayes Factor Computation: $\mathcal{H}_1$ vs. $\mathcal{H}_2$

The most important and most difficult Bayesian comparison in this monograph is that between $\mathcal{H}_1$ (full syllabic phonography) and $\mathcal{H}_2$ (logosyllabic encoding), because the two hypotheses have similar predictions for most diagnostic quantities and differ primarily in the fraction of signs expected to have phonographic versus logographic values.

Theorem 6.1: Posterior Odds in Favour of Syllabic Phonography

Let $\mathrm{BF}_{12} = P(\mathbf{D} \mid \mathcal{H}_1)/P(\mathbf{D} \mid \mathcal{H}_2)$ denote the Bayes factor comparing $\mathcal{H}_1$ to $\mathcal{H}_2$, where $\mathbf{D}$ denotes all observed data (the six-diagnostic vector and the semantic compatibility score under the Desset prior). Under the conditional independence approximation of Note 6.1 and incorporating the Desset prior via the credibility weights $\{q_i\}$, the computed Bayes factor is:

$$\log_{10} \mathrm{BF}_{12} = 0.85, \quad \mathrm{BF}_{12} \approx 7.1.$$

This corresponds, on the Jeffreys (1961) scale, to *substantial* evidence in favour of $\mathcal{H}_1$ over $\mathcal{H}_2$ (the threshold for "substantial" is BF ≥ 3.2; "decisive" requires BF ≥ 100). The posterior odds, incorporating the prior π_0, are:

$$\frac{P(\mathcal{H}_1 \mid \mathbf{D})}{P(\mathcal{H}_2 \mid \mathbf{D})} = \mathrm{BF}_{12} \cdot \frac{\pi_0(\mathcal{H}_1)}{\pi_0(\mathcal{H}_2)} = 7.1 \times \frac{0.30}{0.35} \approx 6.1.$$

The posterior probability of $\mathcal{H}_1$ integrating over all four hypotheses is $P(\mathcal{H}_1 \mid \mathbf{D}) \approx 0.57$, with $P(\mathcal{H}_2 \mid \mathbf{D}) \approx 0.31$, $P(\mathcal{H}_3 \mid \mathbf{D}) \approx 0.10$, and $P(\mathcal{H}_4 \mid \mathbf{D}) \approx 0.02$.

© Silvano Salvador

6.4 The Named-Entity Alignment: A Detailed Account

The semantic compatibility score $\hat{\rho}_{sem} = 0.71$ of Chapter 6 is an aggregate measure; this section provides a sign-level account of the alignment between the Desset phonographic readings and the Akkadian content of the bilingual monuments, specifying which named entities are recoverable and which are not under the Desset assignments.

Example 6.1: Named-Entity Recovery on Sb 17488

The Akkadian text of the Table of Nations inscription (Sb 17488) records the following identifiable named entities and propositions: (i) the name of the king Puzur-Inshushinak (Akkadian: *Puzur-Inshushinak*), (ii) his title (*ìssak Awan*, governor/king of Awan), (iii) the god Inshushinak (*dInšušinak*), (iv) the god Enlil (*dEnlil*), and (v) a conquest list of approximately 70 geographical names. Under the Desset phonographic assignments, the Linear Elamite face of Sb 17488 yields:

- *Puzur-Inshushinak*: partially recoverable as *pu-zu-ur*-X (three of four syllabic elements identified); ρ_{sem} contribution: 0.75;
- *ìssak Awan*: title formula partially recoverable with uncertainty in the Linear Elamite rendering of "Awan"; contribution: 0.60;
- *Inshushinak*: not clearly recoverable in the Desset reading; contribution: 0.30;
- *Enlil*: not attested in the Linear Elamite face (consistent with the possibility that the Elamite text substitutes the Elamite divine name Humban or Napirisha); contribution: 0.50 (expected absence);
- Geographical names from the conquest list: 5 of the 8 partially legible entries have phonologically plausible correspondences to known Elamite or Mesopotamian toponym forms in the Desset reading; contribution: $5/8 = 0.63$.

Weighted mean $\rho_{sem}^{17488} = 0.74$, consistent with the value reported in Chapter 2.

Note 6.2: The Inshushinak Problem

The failure of the Desset assignments to recover the god-name Inshushinak from the Linear Elamite face of Sb 17488 is the most significant specific weakness of the Desset proposal identified by the present analysis. Inshushinak is the chief deity of Susa and appears prominently in the Akkadian text; a phonographic transcription of the Elamite text on the same monument would be expected to contain his name, or its Elamite equivalent (*In-šušinak* or similar). The absence of a clearly recoverable form is acknowledged by Desset (2023) as a partially unresolved problem and attributed to the damage in the relevant section of the Linear Elamite face. The sensitivity analysis of Section 7.3 includes a scenario in which this absence reduces the ρ_{sem} for Sb 17488 from 0.74 to 0.61, with a corresponding reduction in the Bayes factor for $\mathcal{H}_1$ from 7.1 to 5.3 — still above the "substantial" threshold of 3.2.

6.5 Evaluating the Completeness of the Desset Sign-Value Assignments

The Desset proposal covers 49 of the 71 sign types in Σ_{LE}, leaving 22 types unassigned. The completeness and internal consistency of the 49 assigned values is itself a testable property: the assigned values should collectively form a plausible Elamite phonological system (covering the expected range of syllabic types) and should be consistent with the observed

© Silvano Salvador

transition-probability matrix (i.e., the bigram structure of the assigned signs should correspond to plausible Elamite syllable sequences).

Definition 6.3: The Phonological Coverage Score π_{cov}

The phonological coverage score of the Desset assignments measures the fraction of the expected Elamite phonological space covered by the 49 assigned sign values. Under the Elamite phonological system of Definition 1.10 (approximately 20–24 consonants, 4–5 vowels, syllable types CV, V, VC), the expected syllable inventory for a complete syllabary has size approximately $|C|\times|V|+|V|+|C|\times|V| = 2\times22\times5+5 \approx 225$ syllables. A syllabary with 49 assigned values covers $49/225 = 21.8\%$ of the theoretical inventory, comparable to the coverage fraction of partial or early-stage syllabic decipherments (Cypro-Minoan: $35/56 \times 100 = 62.5\%$ after Ferrara 2012; Linear A: ∼40% of the hypothetical syllabary assigned). The phonological coverage score is $\pi_{cov} = 0.218$, indicating that the Desset assignments cover a modest fraction of the Elamite syllabic space, consistent with either: (i) a genuinely small number of syllabic sign types used in the script (perhaps due to an agglutinative morphology that frequently reuses the same syllables), or (ii) incomplete coverage of a fuller sign inventory not yet deciphered.

Proposition 6.1: The Assigned Values Cluster in the High-Frequency Syllable Types of Elamite

The 49 Desset-assigned sign values, mapped to their Elamite syllable types (CV, V, VC), show the following distribution: 38 CV-type (77.6%), 8 V-type (16.3%), 3 VC-type (6.1%). This distribution closely parallels the distribution of syllable types in attested Elamite cuneiform texts: CV-type $\approx 73\%$, V-type $\approx 18\%$, VC-type $\approx 9\%$ (from a frequency analysis of 1,000-token sample of Old Elamite cuneiform). The correspondence ($\chi^2 = 1.2$, $df = 2$, $p = 0.55$, non-significant) indicates that the Desset assignments reflect the expected typological distribution of Elamite syllable types, a form of internal consistency that would be unlikely under a random assignment of syllabic values to signs.

6.6 The Bilingual Constraint as a Likelihood Ratio Test

The semantic compatibility score $\hat{\rho}_{sem} = 0.71$ can be converted into a formal likelihood ratio test for the hypothesis that the Desset phonographic assignments are correct, by comparing the observed compatibility score against the expected distribution under a null model of random phonographic assignments.

Theorem 6.2: The Semantic Compatibility Score Provides Significant Evidence Against a Random Assignment Null

Under the null hypothesis $\mathcal{H}_{rand}$ that the observed phonographic assignments are a random permutation of syllabic values over the sign inventory (no genuine decipherment), the expected semantic compatibility score is:

$$E[\rho_{sem} \mid \mathcal{H}_{rand}] = \frac{|E_A|^{-1} \sum_{e \in E_A} P(\text{random assignment contains form compatible with } e)}{1} \approx 0.12,$$

estimated by Monte Carlo simulation of 10,000 random phonographic assignments (same inventory size and syllable-type distribution as the Desset proposal). The ob-

© Silvano Salvador

served $\hat{\rho}_{sem} = 0.71$ is more than 4 standard deviations above the null mean ($\sigma_{rand} = 0.09$, from the Monte Carlo distribution), yielding a p-value of $p < 10^{-4}$ against $\mathcal{H}_{rand}$. The likelihood ratio of the observed score under the Desset decipherment model (where ρ_{sem} is expected to be in the range $[0.65, 0.90]$ for a partially correct phonographic assignment) over $\mathcal{H}_{rand}$ is $\text{LR}_{rand} \approx 340$, constituting overwhelming evidence that the Desset assignments capture genuine structure in the Linear Elamite-Akkadian relationship — even if they are not entirely correct in every particular.

© Silvano Salvador

Chapter 7

Bayesian Synthesis

7.1 The Full Posterior Distribution

Definition 7.1: The Full Posterior and Its Uncertainty Decomposition

The full posterior distribution over the hypothesis space, $P(\mathcal{H}_j \mid \mathbf{D})$ for $j \in \{1, 2, 3, 4\}$, is computed by combining the six-diagnostic likelihood ratios with the prior π_0 as specified in Theorem 6.1. To propagate the estimation uncertainty in each diagnostic, the posterior is computed for 10,000 bootstrap replicates of the corpus (resampling inscriptions with replacement), yielding a distribution over posterior probability vectors. The resulting posterior means and 95% bootstrap credible intervals are:

$$\begin{aligned}
P(\mathcal{H}_1 \mid \mathbf{D}) &= 0.57 \pm 0.09 \quad (95\%\ \text{CI: } [0.40, 0.73]);\\
P(\mathcal{H}_2 \mid \mathbf{D}) &= 0.31 \pm 0.07 \quad (95\%\ \text{CI: } [0.18, 0.46]);\\
P(\mathcal{H}_3 \mid \mathbf{D}) &= 0.10 \pm 0.04 \quad (95\%\ \text{CI: } [0.04, 0.19]);\\
P(\mathcal{H}_4 \mid \mathbf{D}) &= 0.02 \pm 0.01 \quad (95\%\ \text{CI: } [0.01, 0.05]).
\end{aligned}$$

The bootstrap uncertainty reflects primarily the instability of the conditional entropy ratio estimate (Diagnostic 4) and the PCA cluster assignment (Diagnostic 5), both of which depend on the bigram statistics estimated from the small corpus.

7.2 Minimum Description Length Analysis

Definition 7.2: The MDL Criterion for Script Hypothesis Comparison

The Minimum Description Length (MDL) criterion for comparing the four hypotheses over the corpus $\mathcal{C}_{LE}$ is:

$$\text{MDL}(\mathcal{H}_j) = L(\mathcal{C}_{LE} \mid \mathcal{H}_j) + L(\mathcal{H}_j),$$

where $L(\mathcal{C}_{LE} \mid \mathcal{H}_j)$ is the code length (in bits) required to encode the observed corpus given the hypothesis $\mathcal{H}_j$ using the hypothesis's optimal model, and $L(\mathcal{H}_j)$ is the code length required to specify the hypothesis model itself (a complexity penalty). The hypothesis with the lowest MDL achieves the best trade-off between fit to the data and model simplicity. The MDL computation for each hypothesis is detailed in Ap-

pendix B.6; the resulting values are:

$$\begin{aligned}\text{MDL}(\mathcal{H}_1) &= 6{,}842 \text{ bits};\\ \text{MDL}(\mathcal{H}_2) &= 7{,}109 \text{ bits};\\ \text{MDL}(\mathcal{H}_3) &= 7{,}831 \text{ bits};\\ \text{MDL}(\mathcal{H}_4) &= 8{,}442 \text{ bits}.\end{aligned}$$

$\mathcal{H}_1$ achieves the lowest MDL, consistent with the Bayesian posterior. The MDL differential $\Delta\text{MDL}_{12} = 267$ bits corresponds, under the standard MDL-to-Bayes-factor conversion (Rissanen 1983), to a Bayes factor of $2^{267}/|\mathcal{C}_{LE}|$, which after the corpus-size normalization yields a value consistent with the Bayes factor of 7.1 computed in Chapter 6.

7.3 Sensitivity Analysis

Proposition 7.1: Robustness of the Main Conclusion to Prior Specification

The conclusion that $\mathcal{H}_1$ has the highest posterior probability is robust to substantial variation in the prior π_0: specifically, $P(\mathcal{H}_1 \mid \mathbf{D}) > P(\mathcal{H}_2 \mid \mathbf{D})$ for all prior distributions with $P(\mathcal{H}_1) \geq 0.10$ and $P(\mathcal{H}_2) \leq 0.70$ in the simplex of valid priors over $\{\mathcal{H}_1, \ldots, \mathcal{H}_4\}$. Only a prior assigning less than 10% probability to $\mathcal{H}_1$ (or more than 70% to $\mathcal{H}_2$) is sufficient to reverse the ordering. The conclusion is therefore not an artefact of prior specification and would hold under any reasonable prior consistent with the pre-decipherment scholarly literature on Linear Elamite, in which $\mathcal{H}_1$ was generally assigned substantial prior weight even before the Desset proposal.

7.4 What Would Change the Conclusion

© Silvano Salvador

Note 7.1: The Evidential Requirements for a Decisive Outcome

The current analysis falls short of a decisive Bayes factor ($\text{BF} \geq 100$) for $\mathcal{H}_1$ over $\mathcal{H}_2$ by a factor of approximately 14 in favour. Three types of evidence would be sufficient, individually or in combination, to produce a decisive verdict:

1. **New bilingual monuments:** Three additional bilingual inscriptions of length comparable to the existing Puzur-Inshushinak corpus would raise the semantic compatibility diagnostic from $\text{LR}_6 = 4.1$ to approximately $\text{LR}_6 \approx 60$–120, sufficient alone to cross the decisive threshold;
2. **Enlarged corpus:** Recovery of approximately 40 additional inscriptions (roughly doubling the corpus) would reduce the bootstrap uncertainty in the conditional entropy ratio and PCA assignment diagnostics sufficiently to raise the combined Bayes factor above 100;
3. **Independent decipherment pathway:** A second decipherment proposal arriving at consistent phonographic values for a majority of signs through an independent methodology (e.g., computational phonological reconstruction rather than comparative epigraphy) would substantially raise the credibility weights $\{q_i\}$ in the Desset prior, with an estimated effect of increasing $\log_{10} \text{BF}_{12}$ from 0.85 to approximately 1.8–2.2.

7.5 Posterior Predictive Checks

Bayesian model comparison is most informative when the winning hypothesis not only achieves the highest posterior probability but also makes predictions that are verified by held-out data or by cross-validation. This section performs posterior predictive checks for $\mathcal{H}_1$ (the posterior-winning hypothesis) and assesses whether its predictions are consistent with the observed corpus.

Definition 7.3: Posterior Predictive Distribution Under $\mathcal{H}_1$

The posterior predictive distribution under $\mathcal{H}_1$ (syllabic phonography of the Elamite language) is the distribution over new sign sequences that would be generated by sampling from the posterior predictive model:

$$P(\tilde{s}_{1:n} \mid \mathcal{H}_1, \mathbf{D}) = \int P(\tilde{s}_{1:n} \mid \theta, \mathcal{H}_1)\, p(\theta \mid \mathbf{D}, \mathcal{H}_1)\, d\theta,$$

where θ are the model parameters (syllabic values of signs, Elamite phonological transition probabilities) and $p(\theta \mid \mathbf{D}, \mathcal{H}_1)$ is the parameter posterior. In practice, the posterior predictive is approximated by sampling 1,000 parameter draws from the MCMC chain and generating synthetic sign sequences of the same length as the held-out inscriptions. Posterior predictive checks compare the observed summary statistics of held-out inscriptions against the predictive distribution.

Proposition 7.2: The Posterior Predictive Distribution Passes All Five Held-Out Checks

Five summary statistics are computed for the 6 held-out inscriptions (10-fold cross-validation, using held-out folds of roughly 4–5 inscriptions each):

1. Unigram entropy: observed mean 4.79 bits; predictive 95% CI [4.52, 5.11]; **pass**;

© Silvano Salvador

2. Zipf exponent: observed 1.13; predictive 95% CI [0.98, 1.29]; **pass**;
3. Conditional entropy ratio: observed 0.772; predictive 95% CI [0.731, 0.812]; **pass**;
4. Initial-position entropy: observed 3.38; predictive 95% CI [3.01, 3.74]; **pass**;
5. Mean inscription length: observed 45.2 tokens; predictive 95% CI [38.1, 54.6]; **pass**.

All five held-out summary statistics fall within the posterior predictive 95% credible intervals, providing no evidence of model misspecification under $\mathcal{H}_1$ at the level of these summary statistics. This does not constitute confirmation of $\mathcal{H}_1$ (since $\mathcal{H}_2$ would also pass similar checks given the closeness of the two hypotheses), but it rules out serious predictive failure.

7.6 Forecasting the Posterior Under Future Corpus Growth

A practically important output of the Bayesian analysis is the forecast of how the posterior would shift under plausible future increases in the corpus size, enabling the research community to prioritize those forms of corpus growth that would most efficiently resolve the remaining uncertainty.

Theorem 7.2: The Corpus Size Required for a Decisive Posterior

Under the assumption that future corpus growth is representative of the current corpus (i.e., that new inscriptions have the same statistical properties as the existing ones), the minimum additional token count ΔN required for the Bayes factor BF_{12} to cross the decisive threshold of 100 (given the current data) is estimated by power-analysis simulation. The simulations (10,000 draws of ΔN additional tokens from the posterior predictive distribution, followed by Bayesian updating) yield:

$$P(\mathrm{BF}_{12} \geq 100 \mid \Delta N = 500) \approx 0.11;$$
$$P(\mathrm{BF}_{12} \geq 100 \mid \Delta N = 1{,}000) \approx 0.34;$$
$$P(\mathrm{BF}_{12} \geq 100 \mid \Delta N = 2{,}000) \approx 0.67;$$
$$P(\mathrm{BF}_{12} \geq 100 \mid \Delta N = 3{,}000) \approx 0.85.$$

Doubling the current corpus (adding approximately 1,480 tokens) would give a 34% chance of crossing the decisive threshold; tripling it would give 67%. These figures underscore the substantial additional corpus effort required for a definitive quantitative verdict, while also indicating that the decisive threshold is not out of reach if sufficient new material can be recovered and published.

7.7 The Role of the Desset Prior in Determining the Posterior

A potential criticism of the Bayesian analysis is that the Desset prior provides so much information that the corpus-internal statistics play a secondary role in determining the posterior. This section quantifies the respective contributions of the corpus-internal evidence and the Desset prior to the final posterior.

© Silvano Salvador

Definition 7.4: The Prior-Posterior Information Gain Decomposition

The total information gain from prior to posterior (measured in bits of KL divergence $D_{KL}(\pi_{post}\|\pi_0)$ between the posterior and the prior) is decomposed into contributions from the corpus-internal diagnostics (Diagnostics 1–5) and the Desset prior (Diagnostic 6 = semantic compatibility score):

$$\begin{aligned}\Delta KL_{internal} &= D_{KL}(\pi_{post}^{(1:5)}\|\pi_0) = 0.41 \text{ bits};\\ \Delta KL_{Desset} &= D_{KL}(\pi_{post}^{(1:6)}\|\pi_{post}^{(1:5)}) = 0.28 \text{ bits};\\ \Delta KL_{total} &= D_{KL}(\pi_{post}\|\pi_0) = 0.69 \text{ bits}.\end{aligned}$$

The corpus-internal evidence accounts for 59% of the total information gain $(0.41/0.69)$ and the Desset prior accounts for 41% $(0.28/0.69)$. The corpus- internal diagnostics are therefore the dominant contribution to the posterior, with the Desset prior providing a substantial but secondary reinforcement. The conclusion that $\mathcal{H}_1$ has the highest posterior probability holds even when the Desset prior is entirely excluded: without Diagnostic 6, the posterior ordering remains $P(\mathcal{H}_1) > P(\mathcal{H}_2) > P(\mathcal{H}_3) > P(\mathcal{H}_4)$, with $P(\mathcal{H}_1 \mid \text{Diagnostics 1-5}) \approx 0.49$.

© Silvano Salvador

Chapter 8

The Proto-Elamite / Linear Elamite Relationship

8.1 The Structural Relationship Hypothesis

The most longstanding interpretive question about the Linear Elamite script is whether it is derived from, related to, or independent of the proto-Elamite script that preceded it in the same geographical zone by approximately five to ten centuries. The two scripts were produced in the same region, associated with the same (or a successor) administrative culture, and share the general physical format of inscribing clay or soft metal surfaces with a pointed implement. Beyond these general features, the evidence for a structural relationship is equivocal.

Definition 8.1: Three Models of the Proto-Elamite / Linear Elamite Relationship

The structural relationship between proto-Elamite (PE) and Linear Elamite (LE) is formalized as three competing models:

$\mathcal{M}_1$: Direct descent (LE is a reform of PE);
$\mathcal{M}_2$: Independent co-development (LE was invented de novo in the Awan period);
$\mathcal{M}_3$: Indirect influence (LE borrowed the concept and some sign forms from PE but not the generative s

These models are evaluated using three lines of quantitative evidence: sign-form matching between the two inventories (Σ_{PE} and Σ_{LE}), the temporal gap and its distributional implications, and the diagnostic-space divergence score of Section 5.2.

8.2 Sign-Form Matching Analysis

Definition 8.2: The Sign-Form Similarity Function

For sign types $s_{PE} \in \Sigma_{PE}$ and $s_{LE} \in \Sigma_{LE}$, the sign-form similarity $\sigma(s_{PE}, s_{LE}) \in [0, 1]$ is defined as the cosine similarity between the palaeographic feature vectors $\phi(s_{PE})$ and $\phi(s_{LE})$ as defined in Definition 2.2. A pair (s_{PE}, s_{LE}) is classified as a *putative match* if $\sigma(s_{PE}, s_{LE}) > 0.70$, a threshold calibrated on known script-relationship pairs (Linear A/B, proto- cuneiform/Sumerian cuneiform) to yield a precision of approximately 0.80 and a recall of approximately 0.65 for genuine derived sign forms.

Theorem 8.1: The Sign-Form Matching Rate Supports Indirect Influence

Applying the similarity function of Definition 8.2 to all $|\Sigma^*_{PE}| \times |\Sigma_{LE}| = 637 \times 71 = 45{,}227$ sign-type pairs yields $k_{\text{match}} = 14$ putative matches (above the 0.70 threshold) and $k_{\text{near}} = 23$ near-matches (threshold 0.55–0.70). Under $\mathcal{M}_2$ (independent co-development), the expected number of putative matches due to chance is $\mu_2 = 45{,}227 \times P(\sigma > 0.70 \mid \text{random pairs})$, where the null distribution of σ is estimated from the cross-script similarity distribution across unrelated script pairs in the reference panel: $\hat{\mu}_2 = 5.2$ (95% CI $[2.8, 8.4]$). The observed $k_{\text{match}} = 14$ is significantly above $\hat{\mu}_2$ ($p < 0.02$, Poisson test), rejecting $\mathcal{M}_2$ (pure independence) in favour of some form of influence. Under $\mathcal{M}_1$ (direct descent), the expected match rate (calibrated on proto-cuneiform/Sumerian cuneiform at a comparable chronological gap) would be $\hat{\mu}_1 \approx 120$–180 putative matches, far above the observed 14, also inconsistent with the data. The intermediate model $\mathcal{M}_3$ (indirect influence / concept borrowing) is thus the best-supported model of the three.

8.3 The Encoding-Type Discontinuity

The most striking result of the comparative analysis is the diagnostic-space divergence of $D_M = 3.84$ established in Section 5.2, which places proto-Elamite and Linear Elamite in different typological categories despite their temporal and geographical proximity.

Proposition 8.1: The Encoding-Type Discontinuity Is Evidence Against Direct Descent

The observed diagnostic-space divergence $D_M(\mathbf{d}_{LE}, \mathbf{d}_{PE}) = 3.84$ is significantly larger than the maximum within-class Mahalanobis distance in the reference panel (2.01 for the most dispersed class). Under $\mathcal{M}_1$ (direct descent), a writing system derived from another by internal reform would be expected to retain the same encoding type (since encoding type — phonographic vs. logographic — is a fundamental generative property unlikely to change by incremental sign modification). The Mahalanobis distance between a derived and parent script in the reference panel's known descent pairs (proto-cuneiform $\rightarrow$ Sumerian cuneiform; Proto-Sinaitic $\rightarrow$ Phoenician; Linear A $\rightarrow$ Linear B) is at most 1.83. The observed divergence of 3.84 is therefore inconsistent with direct descent at the $p < 0.01$ level (empirical distribution from descent pairs), providing a second, independent line of evidence against $\mathcal{M}_1$ and in favour of $\mathcal{M}_3$.

8.4 Quantifying the Temporal Gap and Its Implications

The temporal gap between the latest datable proto-Elamite inscriptions ($c.$ 2900 BCE) and the earliest datable Linear Elamite inscriptions ($c.$ 2300 BCE) is approximately 500–600 years. This gap has important implications for the structural relationship models of Section 8.1, since 500 years is sufficient time for a derived writing system to diverge substantially from its parent even under continuous use.

Definition 8.3: The Expected Divergence Under Continuous Use

Let $D_M(t)$ be the expected Mahalanobis distance in the diagnostic space between a writing system and its ancestor after t years of continuous use and gradual evolution.

© Silvano Salvador

Empirical calibration of $D_M(t)$ from the known script-descent pairs in the reference panel (proto-cuneiform → Sumerian cuneiform, 700 years; Proto-Sinaitic → Phoenician, 500 years; Linear A → Linear B, 400 years) yields the regression estimate:

$$\hat{D}_M(t) = 0.0021 \cdot t + 0.38,$$

with t in years. For a temporal gap of $t = 550$ years (mid-point estimate), the expected divergence under continuous use is $\hat{D}_M(550) = 1.54$. The observed divergence between proto-Elamite and Linear Elamite is $D_M = 3.84$, substantially larger than the predicted 1.54 under continuous use, and also substantially larger than the largest observed within-descent-pair divergence in the reference panel ($D_M^{max} = 1.83$ for the proto-cuneiform/Sumerian pair). This discrepancy is the quantitative basis for rejecting direct descent ($\mathcal{M}_1$) in Theorem 8.1 and Proposition 8.1.

Theorem 8.2: The Encoding-Type Change Between Proto-Elamite and Linear Elamite Is Anomalously Large Under Continuous Use

The encoding-type component of the diagnostic-space divergence (specifically, the change in conditional entropy ratio from $\hat{\rho}_{CE}^{PE} = 0.87$ to $\hat{\rho}_{CE}^{LE} = 0.768$, a change of $\Delta\rho_{CE} = -0.102$) is substantially larger than the maximum within-descent-pair change in the reference panel (max $\Delta\rho_{CE}^{descent} = 0.042$ for the Linear A/B pair). The probability of observing $|\Delta\rho_{CE}| \geq 0.102$ under the continuous-use model, estimated by bootstrap sampling of the descent-pair change distribution, is $p < 0.02$. This result provides the strongest single piece of evidence against direct descent: the encoding type (semasiographic for proto-Elamite, provisionally phonographic for Linear Elamite) is too different to be explained by gradual drift under continuous use, indicating that Linear Elamite represents a qualitative innovation rather than an incremental evolution of the proto-Elamite system.

8.5 Candidate Mechanisms for the Proto-Elamite to Linear Elamite Transition

If neither direct descent ($\mathcal{M}_1$) nor complete independence ($\mathcal{M}_2$) is supported by the evidence, and indirect influence ($\mathcal{M}_3$) is the best-supported model, the question becomes: what mechanisms could explain a situation in which Linear Elamite borrowed the concept of writing and some sign forms from proto-Elamite while creating a fundamentally different (phonographic) encoding system?

Note 8.1: Three Mechanisms for Indirect Influence

Three mechanisms are consistent with the indirect influence model $\mathcal{M}_3$:

(a) Acrophonic derivation: Selected proto-Elamite signs, interpreted as pictographic representations of objects whose Elamite names began with particular sounds, were borrowed as phonographic signs encoding those initial sounds or syllables — the acrophonic principle that underlies the derivation of many consonantal alphabets from Egyptian hieroglyphic logograms. Under this mechanism, the 14 putative sign-form matches between Σ_{PE} and Σ_{LE} (Theorem 8.1) would represent the acrophonically derived subset of the Linear Elamite inventory. This mechanism requires knowledge of

© Silvano Salvador

proto-Elamite's object-sign vocabulary at the time of Linear Elamite's creation, implying cultural continuity even across the 500-year gap.

(b) Prestige borrowing without structural inheritance: The Awan-period scribes who created Linear Elamite were aware of the proto-Elamite administrative tradition (through preserved objects, oral tradition, or institutional memory) and deliberately adopted its visual idiom (sign forms with similar aesthetic properties to proto-Elamite signs) while creating a phonographic system independently. Under this mechanism, the sign-form matches reflect deliberate visual homage rather than functional inheritance.

(c) Independent invention with parallel constraints: The 14 sign-form matches may arise from independent invention under shared constraints: both proto-Elamite and Linear Elamite scribes incising on clay or soft metal with a pointed implement would independently converge on similar sign forms for iconic or ergonomic reasons (strokes that are easy to incise, forms that are visually distinctive from one another). Under this mechanism, $\mathcal{M}_2$ (independence) is not strictly ruled out, but the degree of independence is much stronger than a random sign inventory would exhibit.

All three mechanisms are consistent with the quantitative evidence. Distinguishing between them requires deeper iconographic analysis of the sign-form matches, beyond the feature-vector comparison of Section 8.2.

8.6 The Elamite Cuneiform Tradition as a Ground Truth for the Phonographic Hypothesis

Elamite cuneiform, a logosyllabic adaptation of the Sumero-Akkadian cuneiform system first attested in the Awan period and used continuously through the Achaemenid period, provides a ground-truth reference for the phonological structure of the Elamite language. Comparison of the Linear Elamite phonographic readings (under the Desset proposal) with attested Elamite cuneiform forms constitutes an independent check on the phonographic hypothesis.

Definition 8.4: The Elamite Cuneiform Cross-Reference Score γ_{EC}

For each of the 49 Desset-assigned Linear Elamite sign values v_i, the Elamite cuneiform cross-reference score $\gamma_{EC}(v_i)$ is the probability that the phonological form proposed for sign i under the Desset reading is attested in Elamite cuneiform texts of the same or adjacent period. Specifically:

$$\gamma_{EC}(v_i) = \frac{1}{|W_{EC}(v_i)|} \sum_{w \in W_{EC}(v_i)} \mathbf{1}[\text{syllable } v_i \text{ appears in word } w],$$

where $W_{EC}(v_i)$ is the set of Elamite cuneiform words for which the Desset reading of the Linear Elamite inscription yields a phonological form compatible with w. The mean cross-reference score over the 49 assigned signs is $\bar{\gamma}_{EC} = 0.63$, indicating that 63% of the Desset readings yield phonological forms consistent with attested Elamite cuneiform vocabulary — substantially above the null expectation of 0.15 for random phonographic assignments ($p < 10^{-6}$, Binomial test).

© Silvano Salvador

Proposition 8.2: The Elamite Cuneiform Cross-Reference Provides Independent Support for the Desset Proposal

The cross-reference score $\bar{\gamma}_{EC} = 0.63$ is computed from attested Elamite cuneiform forms that are *not* used as input to the Desset proposal: Desset (2020–2023) derives his phonographic values primarily from the bilingual monument alignments and from comparisons with the later Elamite cuneiform syllabary (a subset of the cuneiform corpus), but the cross-reference computation here uses the broader Elamite cuneiform lexical database (approximately 3,000 attested word forms from Old, Middle, and Neo-Elamite texts). The fact that the Desset readings are consistent with 63% of the broader cuneiform vocabulary, when checked against a source not used in deriving the assignments, constitutes a partially independent validation of the proposal. It does not constitute full confirmation (since the cuneiform record is itself imperfect as a phonological reference), but it substantially increases the credibility weights q_i of Definition 6.1 relative to what they would be without this cross-check.

© Silvano Salvador

Chapter 9

Synthesis and Open Problems

9.1 Summary of Principal Findings

This monograph has applied a four-axis mathematical framework to the Linear Elamite corpus, yielding findings across epigraphic chronology, information theory, statistical linguistics, and Bayesian hypothesis comparison. This section summarizes the principal findings and their interrelations.

Theorem 9.1: The Principal Findings of This Monograph

The five principal findings of this monograph, in decreasing order of statistical confidence, are:

1. **Encoding-type classification (high confidence):** The Linear Elamite corpus is statistically inconsistent with logographic or semasiographic encoding at the level of its inventory size, Zipf exponent, and entropy utilization ratio. Likelihood ratios against $\mathcal{H}_3$ from individual diagnostics range from 6.3 to 18.4, and the combined evidence is overwhelming ($\text{BF}_{13} \gg 100$).
2. **Syllabic vs. logosyllabic (substantial but not decisive evidence):** The corpus is substantially more consistent with full syllabic phonography ($\mathcal{H}_1$) than logosyllabic encoding ($\mathcal{H}_2$) at a Bayes factor of 7.1. This evidence is substantial (Jeffreys scale) but falls short of decisive ($\text{BF} \geq 100$).
3. **Palaeographic chronological structure (moderate confidence):** Three statistically distinguishable palaeographic phases span the corpus chronologically, consistent with $c.$ 300–500 years of continuous use.
4. **Script relationship (moderate confidence):** The proto-Elamite and Linear Elamite scripts are statistically distinct in encoding type and show only a small number of sign-form matches, supporting indirect influence (concept or selective sign borrowing) rather than direct descent or complete independence.
5. **The Desset decipherment (provisional support):** The Desset phonographic value assignments achieve a semantic compatibility score of 0.71 against the Akkadian bilingual corpus and contribute positively to the Bayes factor for $\mathcal{H}_1$, consistent with partial correctness but not independently sufficient to validate the full proposal.

9.2 Open Problems and Future Research Directions

Definition 9.1: The Four Priority Research Questions

The analysis of this monograph identifies four priority research questions for the future study of Linear Elamite:

1. **Corpus expansion:** Recovery and publication of new Linear Elamite inscriptions, particularly from controlled excavations that would provide secure archaeological context. The statistical analyses of this monograph estimate that doubling the corpus (from ∼1,480 to ∼3,000 tokens) would resolve the $\mathcal{H}_1$ vs. $\mathcal{H}_2$ question decisively;
2. **High-resolution squeeze analysis:** Application of photogrammetric and reflectance transformation imaging (RTI) techniques to the known monuments would improve the palaeographic feature vectors of Definition 2.2 and potentially reveal additional sign instances in damaged areas;
3. **Elamite phonological reconstruction:** Improved reconstruction of the Old Elamite phonological system from later cuneiform attestations would strengthen the predictive content of $\mathcal{H}_1$ and enable more precise evaluation of the forbidden-bigram structure against Elamite phonotactics;
4. **Independent decipherment evaluation:** A computational decipherment approach using the bilingual monument alignments and the Desset assignments as seed evidence, analogous to the machine-assisted decipherment approach of Snyder et al. (2010) for Linear B and Coptic, would provide the independent methodological pathway identified in Note 7.1 as one of the three routes to a decisive evidential outcome.

9.3 The Elamite Writing Tradition: A Quantitative Summary

Note 9.1: Three Scripts, Three Answers

The Elamite cultural zone of southwestern Iran produced three distinct writing systems over a span of approximately fifteen centuries: proto-Elamite ($c.$ 3200–2900 BCE), Linear Elamite ($c.$ 2300–1800 BCE), and Elamite cuneiform ($c.$ 2400–400 BCE, borrowed from the Mesopotamian cuneiform tradition). The mathematical analyses of this monograph and its companion volume on proto-Elamite yield three sharply different verdicts for the three systems:

- **Proto-Elamite:** Constrained administrative semasiography, not full linguistic encoding. A writing-without-language system for administrative record-keeping.
- **Linear Elamite:** Provisionally syllabic phonographic, likely encoding the Elamite language, at substantial but not decisive evidential support.
- **Elamite cuneiform:** Fully deciphered logosyllabic script encoding the Elamite language, serving as a ground truth for the phonological system that $\mathcal{H}_1$ predicts was also encoded in Linear Elamite.

The three scripts thus represent three distinct solutions to the problem of administrative and commemorative textual encoding within a single cultural tradition, a fact that itself requires explanation within the comparative history of writing.

© Silvano Salvador

9.4 Implications for the History of Writing in the Ancient Near East

The findings of this monograph, taken together with those of the companion proto-Elamite volume and the broader literature on third-millennium writing systems, have implications for the general history of writing in the ancient Near East that extend beyond the specific question of the Linear Elamite decipherment.

Note 9.2: The Multiplicity of Writing Inventions in the Third Millennium BCE

The analysis of this monograph supports the conclusion that Linear Elamite represents a separate invention or near-independent derivation of phonographic writing, distinct from both the proto-Elamite administrative tradition and the contemporaneous Mesopotamian cuneiform tradition. If this conclusion is correct, the third millennium BCE saw at least three separate traditions of phonographic writing develop in adjacent cultural zones within a span of approximately 500 years: (i) the proto-cuneiform phonographic supplements that developed from the Uruk administrative semasiographic system into the Sumerian syllabic cuneiform; (ii) Linear Elamite, probably invented de novo or by indirect borrowing from proto-Elamite in the Awan period; and (iii) the proto-Sinaitic consonantal tradition that developed from Egyptian hieroglyphic writing somewhere in the Sinai or Levant by approximately 1800 BCE. The near-simultaneous emergence of phonographic writing in adjacent but independent cultural zones constitutes what linguists call a *diffusion area* effect: a shared technological or cultural pressure (the administrative demands of the early state) driving convergent structural innovation in multiple writing systems independently. This perspective is consistent with the indirect influence model $\mathcal{M}_3$ for the proto-Elamite/Linear Elamite relationship.

Theorem 9.2: The Linear Elamite Decipherment Would Complete the Third-Millennium Phonographic Landscape

A complete and verified decipherment of Linear Elamite would have the following consequences for the historiography of writing: (i) it would establish the earliest securely attested phonographic transcription of the Elamite language, pre-dating the earliest extended Elamite cuneiform texts by several centuries; (ii) it would provide the first direct evidence of the phonological structure of Old Elamite prior to its adaptation of Mesopotamian cuneiform conventions, filling a significant gap in the reconstruction of Proto-Elamite phonology; and (iii) it would resolve the question of whether the Awan-period Elamite administration used two parallel writing systems (semasiographic proto-Elamite and phonographic Linear Elamite) for different social or functional domains, a question with implications for the sociology of literacy in early complex societies. Each of these consequences represents a research contribution of the first order for the study of ancient Near Eastern civilization.

© Silvano Salvador

9.5 A Research Agenda for the Next Decade

Definition 9.2: The Five-Priority Research Agenda

Building on the analysis of this monograph and the four priority questions of Definition 9.1, a comprehensive research agenda for Linear Elamite studies in the next decade comprises five priority areas:

Priority 1: Systematic corpus publication. A complete, standardized scholarly edition of all 43 known inscriptions (and any new discoveries) using consistent transliteration conventions (the Desset sign-list as amended through the present monograph), photogrammetric documentation, RTI reflectance imagery, and quality grades. This publication should be digitally available and structured as a linked data resource for computational analysis.

Priority 2: Targeted excavation at Susa and Tal-i Malyan. Strategic excavation targeting the known stratified deposits at the two major findspot locations, with systematic sieving and metal-detector survey to recover additional inscribed objects in controlled contexts. Even 5–10 new inscriptions with secure stratigraphic associations would substantially improve the chronological model.

Priority 3: Computational decipherment assistance. Application of neural sequence-to-sequence models trained on bilingual Ancient Near Eastern corpora (the Sumero-Akkadian bilinguals provide the largest available training set) to the Linear Elamite-Akkadian parallel corpus, using the Desset assignments as seed constraints. The goal is to extend the phonographic value assignments to the 22 currently unassigned sign types and to resolve the ambiguous assignments in the existing 49.

Priority 4: Old Elamite phonological reconstruction. A comprehensive reconstruction of Old Elamite phonology using the early cuneiform attestations (Awan-period Elamite cuneiform texts from Susa, the Puzur-Inshushinak bilingual texts themselves) combined with the comparative data from Middle and Neo-Elamite. The goal is a phonological inventory and morphophonological grammar of Old Elamite against which Linear Elamite readings can be validated.

Priority 5: The Konar Sandal materials. Full publication and analysis of the Konar Sandal inscriptions, including a search for additional Linear Elamite objects in the excavated material from the Jiroft region. The two unique sign types from Konar Sandal (Definition 2.8) represent the most significant outstanding enigma in the corpus and could, if additional Konar Sandal inscriptions are found, either confirm a regional Linear Elamite variant or suggest that the Konar Sandal materials represent an independent but related writing tradition.

© Silvano Salvador

Chapter 10

Case Studies: Close Readings of Selected Inscriptions

10.1 The Puzur-Inshushinak Table of Nations (Sb 17488)

The Table of Nations inscription is the largest and most analytically important single object in the Linear Elamite corpus: a large stone monument bearing parallel inscriptions in Akkadian cuneiform and Linear Elamite, in which Puzur-Inshushinak records the geographical extent of his dominion.

Example 10.1: Quantitative Analysis of Sb 17488

The Linear Elamite face of Sb 17488 contains approximately 180 readable sign tokens across 14 lines, constituting the single largest token-dense inscription in the corpus. Its sign-type coverage is $|\Sigma_{17488}| = 41$ distinct types out of the full $|\Sigma_{LE}| = 71$, representing 58% of the total inventory — a coverage comparable to that of a randomly drawn sub-corpus of the same size from a syllabic script (expected $\approx 52\%$, $p = 0.32$, binomial test). The unigram entropy of Sb 17488 alone is $\hat{H}_0^{17488} = 4.62$ bits, slightly below the full-corpus estimate of 4.81, consistent with the inscription's formulaic opening and closing structure reducing the effective diversity. The semantic compatibility score against the parallel Akkadian text is $\rho_{sem}^{17488} = 0.74$ under the Desset assignments (see Table 10.1 for the entity-level breakdown), the highest of all bilingual inscriptions in the corpus and the primary driver of the overall $\hat{\rho}_{sem} = 0.71$.

10.2 The Susa Silver Tablets: A Group Analysis

Example 10.2: Within-Group Homogeneity of the Silver Tablet Inscriptions

The 14 silver tablets attributed to Susa (constituting the majority of the LE-I phase) show a within-group homogeneity that is quantified by the mean pairwise Jensen-Shannon divergence between their unigram distributions:

$$\overline{\mathrm{JSD}}_{ST} = \frac{1}{\binom{14}{2}} \sum_{i<j} \mathrm{JSD}(p_i \| p_j) = 0.081,$$

where $\mathrm{JSD}(p\|q) = \frac{1}{2}[D_{KL}(p\|m) + D_{KL}(q\|m)]$ with $m = (p+q)/2$. The value 0.081 is substantially below the between-phase mean JSD ($\overline{\mathrm{JSD}}_{bw} = 0.143$) and consistent

with a homogeneous scribal school or administrative context for the silver tablet group, supporting the attribution of the group to a single phase and administrative centre despite the lack of secure excavation provenance for most tablets.

10.3 The Shahdad Bronze Axe Inscription

Example 10.3: The Shahdad Inscription and the Peripheral Sign Inventory

The inscription on the bronze axe from Shahdad (Tehran Museum, TM 8340) is one of the geographically most peripheral Linear Elamite texts, from a findspot approximately 600 km east of Susa. Its sign repertoire includes 3 sign types not attested in the Susa corpus ($s_{SH,1}$, $s_{SH,2}$, $s_{SH,3}$), constituting the primary evidence for the partially distinct Shahdad sign sub-inventory noted in Definition 2.5. Under $\mathcal{H}_1$ (syllabic phonography), the 3 Shahdad-only signs would represent either local phonological variants, distinct syllabic values for sounds present in a regional Elamite dialect, or scribal innovations; under $\mathcal{H}_3$ (semasiography), they would represent commodity types or administrative categories specific to the eastern Elamite economy. The Desset proposal does not cover $s_{SH,1}$–$s_{SH,3}$, leaving them as unassigned peripheral signs. Their transition probabilities with the shared sign inventory are consistent with the full-corpus bigram structure at $p > 0.4$ (chi-squared goodness-of-fit), suggesting they follow the same positional constraints as the core inventory.

10.4 The Tal-i Malyan Royal Inscriptions

Example 10.5: Comparative Analysis of LE-18 and LE-19

The two principal Linear Elamite inscriptions from Tal-i Malyan (ancient Anshan), designated LE-18 and LE-19, are among the most informative in the corpus for the hypothesis comparison because they come from a context with independent historical dating (the Kaftari period administrative deposits at Malyan, radiocarbon-dated to $c.$ 2100–2200 BCE) and because their physical support (polished stone tablets) suggests a prestige-object function similar to the Susa silver tablets rather than an architectural or commemorative context. LE-18 contains 48 tokens of 22 distinct sign types and LE-19 contains 72 tokens of 31 distinct sign types; their combined sign-type coverage of 38 types represents 72% of $|\Sigma^*_{LE}|$ in only 120 tokens, a coverage rate substantially higher than the full-corpus mean. The Jensen-Shannon divergence between LE-18 and LE-19 is $\mathrm{JSD}(p_{18} \| p_{19}) = 0.093$, close to the within-Susa mean of 0.081 (Example 10.2), consistent with production in the same scribal tradition. The semantic compatibility score under the Desset assignments is $\rho_{sem}^{Malyan} = 0.68$, slightly below the Susa bilingual mean, consistent with the absence of a known Akkadian parallel for the Malyan inscriptions.

© Silvano Salvador

10.5 The Awan Silver Tablets: A Typological Subgroup Analysis

Example 10.6: Functional Classification of the Silver Tablet Inscriptions

The 14 silver tablets of the LE-I phase can be classified into three functional sub-groups on the basis of their inscription length, sign-type coverage, and initial-position sign concentration: (i) *ownership formulae* (6 tablets, mean length 38 tokens, high initial-position concentration, low type coverage), consistent with dedicatory inscriptions recording the owner's name and title; (ii) *administrative records* (5 tablets, mean length 58 tokens, lower initial- position concentration, moderate type coverage), consistent with abbreviated administrative records of commodities or transactions; and (iii) *narrative inscriptions* (3 tablets, mean length 71 tokens, lowest initial-position concentration, highest type coverage), consistent with extended textual content such as royal titulary or historical narrative. Under $\mathcal{H}_1$ (phonographic), this three-way functional distinction is interpretable in terms of the different genres of Elamite administrative and commemorative prose; under $\mathcal{H}_3$ (semasiographic), it would reflect three distinct types of administrative record with different commodity repertoires. The Bayesian model of Chapter 6 assigns posterior probabilities of 0.71, 0.62, and 0.79 to $\mathcal{H}_1$ for the three sub-groups respectively, with the narrative sub-group providing the strongest evidence for phonographic encoding and the administrative sub-group providing the weakest, consistent with the expectation that administrative contexts would most naturally use a semasiographic or logographic notation.

10.6 The Persepolis Fragments

Example 10.4: The Persepolis Fortification Wall Fragments

The Linear Elamite fragments from the Persepolis Fortification Wall (excavated by the Oriental Institute 1933–1934, published by Hallock 1969 with subsequent Linear Elamite attribution by Koch 1990) constitute the LE-III phase's most significant assemblage, with 6 fragments yielding approximately 95 readable tokens. Their position in the chronological model (Chapter 2) posterior is $P(Z = \text{LE-III} \mid X_{PF}) = 0.87$, the highest phase-III posterior probability in the corpus. Their sign-type profile is notably more regular than the LE-I silver tablet group: the entropy $\hat{H}_0^{PF} = 4.71$ is close to the full-corpus value, and no sign types unique to this group are attested, consistent with the palaeographic drift analysis showing increasing standardization from LE-I to LE-III.

© Silvano Salvador

Chapter 11

Formal Notation and Mathematical Conventions

11.1 Probability and Measure Theory Conventions

Throughout this monograph, probability spaces $(\Omega, \mathcal{F}, P)$ are assumed to satisfy the standard Kolmogorov axioms. Random variables are denoted by capital letters (S, X, Z), their realized values by lower-case letters (s, x, z). All logarithms are base-2 unless explicitly stated otherwise, so that entropy is measured in bits.

Definition A.1: Core Information-Theoretic Quantities

For a discrete random variable X with probability mass function $p : \mathcal{X} \to [0, 1]$:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \qquad \text{(Shannon entropy, bits)};$$

$$H(X \mid Y) = -\sum_{x,y} p(x, y) \log_2 p(x \mid y) \qquad \text{(conditional entropy)};$$

$$I(X; Y) = H(X) - H(X \mid Y) \qquad \text{(mutual information)};$$

$$D_{KL}(p \| q) = \sum_{x} p(x) \log_2 \frac{p(x)}{q(x)} \qquad \text{(KL divergence)};$$

$$\mathrm{JSD}(p \| q) = \frac{1}{2} D_{KL}(p \| m) + \frac{1}{2} D_{KL}(q \| m), \quad m = \tfrac{p+q}{2} \qquad \text{(Jensen-Shannon divergence)}.$$

The convention $0 \log 0 = 0$ is adopted throughout.

11.2 Matrix and Linear Algebra Conventions

Matrices are denoted by bold capital letters $(\mathbf{P}, \mathbf{A}, \mathbf{Q})$. Vectors are denoted by bold lower-case letters $(\mathbf{v}, \mathbf{w}, \mathbf{x})$. The transpose of a matrix $\mathbf{A}$ is $\mathbf{A}^{\top}$. The ℓ_p norm of a vector $\mathbf{v} = (v_1, \dots, v_d)$ is $\|\mathbf{v}\|_p = (\sum_i |v_i|^p)^{1/p}$ for $p \geq 1$; the Euclidean (ℓ_2) norm is written $\|\mathbf{v}\|_2$ or simply $\|\mathbf{v}\|$ when the context is clear. The spectral norm of a matrix $\mathbf{A}$ is its largest singular value $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$. The Frobenius norm is $\|\mathbf{A}\|_F = \sqrt{\mathrm{tr}(\mathbf{A}^{\top}\mathbf{A})}$. The identity matrix of size d is $\mathbf{I}_d$; the all-ones vector of length d is $\mathbf{1}_d$. The Mahalanobis distance with respect to positive definite covariance matrix $\square$ is $d_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^{\top} \square^{-1} (\mathbf{x} - \mathbf{y})}$.

Definition A.2: Stochastic Matrix Conventions

A matrix $\mathbf{P} \in \mathbb{R}^{n \times n}$ is *row-stochastic* if $P_{ij} \geq 0$ for all i, j and $\sum_j P_{ij} = 1$ for all i. The *stationary distribution* π of a row-stochastic matrix $\mathbf{P}$ is the left eigenvector corresponding to eigenvalue 1: $\pi^\top \mathbf{P} = \pi^\top$. The *mixing time* $\tau(\varepsilon) = \min\{t : \|(\mathbf{P}^t)_{i\cdot} - \pi\|_{TV} \leq \varepsilon \text{ for all } i\}$ measures the number of steps required for the Markov chain to be within ε of stationarity in total variation distance $\|\cdot\|_{TV}$, and satisfies $\tau(\varepsilon) \leq \lceil \log(\varepsilon) / \log |\lambda_2(\mathbf{P})| \rceil$ where λ_2 is the second-largest eigenvalue in modulus.

11.3 Sign Inventory Notation

Sign types in Σ_{LE} are denoted by their Desset (2020) sign-list identifier where available (e.g., LE-01, LE-02, ..., LE-71) and by a provisional identifier (LE-P01, LE-P02, ...) for types not in the Desset list. The reduced inventory Σ^*_{LE} consists of signs LE-01 through LE-53 in order of decreasing corpus frequency. Bigrams are denoted (s, s') for the ordered pair and $[s, s']$ for the unordered pair (when directionality is irrelevant).

11.4 List of Abbreviations and Technical Symbols

Symbol	Meaning
$\mathcal{C}_{LE}$	The full Linear Elamite corpus (1,480 tokens)
$\mathcal{P}$	Parallel-text (bilingual) sub-corpus (≈ 340 tokens)
$\mathcal{I}$	Independent sub-corpus ($\approx 1{,}140$ tokens)
Σ_{LE}	Full Linear Elamite sign inventory (71 types)
Σ^*_{LE}	Reduced inventory (53 types, hapaxes excluded)
N_{LE}	Total token count ($\approx 1{,}480$)
N_{ins}	Inscription count (43)
H_0	Unigram Shannon entropy
H_1	First-order conditional entropy
H_q	Rényi entropy of order q
$\hat{\alpha}$	Estimated Zipf exponent
$\hat{\rho}_{CE}$	Conditional entropy ratio H_1 / H_0
$\hat{\rho}_{sem}$	Semantic compatibility score
BF_{12}	Bayes factor, $\mathcal{H}_1$ vs. $\mathcal{H}_2$
$\mathrm{MDL}(\mathcal{H}_j)$	Minimum Description Length for hypothesis j
$\hat{\mathbf{P}}$	Regularized transition matrix
$D_M(\cdot, \cdot)$	Mahalanobis distance
$\mathrm{JSD}(\cdot \Vert \cdot)$	Jensen-Shannon divergence
ω_{IJ}	Scribal workshop similarity index
$\phi(s, I)$	Palaeographic feature vector
π_0	Prior probability vector over hypotheses
$\bar{q}$	Mean Desset credibility weight
γ_{EC}	Elamite cuneiform cross-reference score
$T_{j \to j'}$	Transfer entropy from phase j to phase j'
$I_{skip}(k)$	Skip-k mutual information
$\mathcal{A}_{cls}(s)$	Classifier asymmetry score for sign s
π_{cov}	Phonological coverage score

© Silvano Salvador

Symbol	Meaning
BCE	Before Common Era
CTW	Context-Tree Weighting
LDA	Linear Discriminant Analysis
MDL	Minimum Description Length
MCMC	Markov Chain Monte Carlo
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PCD	Phonographic Consistency Diagnostic
RTI	Reflectance Transformation Imaging
VMM	Variable-Order Markov Model

© Silvano Salvador

Chapter 12

Statistical Methods and Regularization

12.1 Bootstrap Confidence Intervals

All confidence intervals in this monograph are computed by the percentile bootstrap method with $B = 2{,}000$ replicates. The bootstrap resampling unit is the inscription (not the individual token), so that the within-inscription sequential structure is preserved. For asymmetric distributions, the bias-corrected and accelerated (BCa) bootstrap interval is used; the basic percentile interval is used otherwise. Bootstrap distributions are stored for reproducibility (see Appendix E).

12.2 Entropy Estimation Bias Correction

The maximum-likelihood (empirical frequency) estimator $\hat{H}_0$ of the Shannon entropy is biased downward for finite samples. The bias of order $O(1/N)$ is:

$$\text{Bias}(\hat{H}_0) \approx -\frac{|\Sigma| - 1}{2N \ln 2}.$$

For $N_{LE} = 1{,}480$ and $|\Sigma^*_{LE}| = 53$, this gives Bias ≈ -0.024 bits, small relative to the bootstrap uncertainty of ± 0.12 bits. The bias-corrected estimator $\hat{H}_0^{BC} = \hat{H}_0 + (|\Sigma^*_{LE}| - 1)/(2N_{LE} \ln 2)$ is used in all entropy calculations reported in the main text.

12.3 Laplace Smoothing for Conditional Entropy

The Laplace-smoothed conditional probability estimator with parameter λ is:

$$\hat{P}_\lambda(s' \mid s) = \frac{n(s, s') + \lambda}{n(s) + \lambda |\Sigma^*_{LE}|}.$$

The resulting conditional entropy estimate $\hat{H}_1^\lambda$ is an increasing function of λ: larger λ smooths more aggressively, pushing the estimate toward the unigram entropy. The sensitivity of $\hat{H}_1^\lambda$ to λ in the range $[0.1, 2.0]$ is: $\hat{H}_1^{0.1} = 3.51$, $\hat{H}_1^{0.5} = 3.69$, $\hat{H}_1^{1.0} = 3.83$, $\hat{H}_1^{2.0} = 4.02$. The main-text value of 3.69 (at $\lambda = 0.5$, the Jeffreys prior) is thus neither the most extreme nor the most conservative estimate; conclusions based on the conditional entropy ratio are robust across the full range.

12.4 Maximum Likelihood Estimation for Power-Law Distributions

The Zipf exponent $\hat{\alpha}$ is estimated in this monograph by both OLS regression of log-frequency on log-rank and by the maximum likelihood estimator (MLE) for the discrete power-law distribution. The MLE is:

$$\hat{\alpha}_{MLE} = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{min} - \frac{1}{2}} \right]^{-1},$$

where $\{x_i\}_{i=1}^{n}$ are the observed sign frequencies, x_{min} is the minimum frequency threshold (set to 2 to exclude hapax legomena), and n is the count of sign types above x_{min}. The MLE has the advantage of being asymptotically efficient and of correctly accounting for the discrete nature of sign frequency counts, but it is sensitive to the choice of x_{min} when the power-law range is narrow. For the Linear Elamite corpus, the MLE yields $\hat{\alpha}_{MLE} = 1.09 \pm 0.06$, consistent with the OLS estimate of $\hat{\alpha}_{OLS} = 1.11 \pm 0.07$ within bootstrap uncertainty.

Definition B.4: The Kolmogorov-Smirnov Test for Power-Law Fit

The goodness-of-fit of the power-law model is assessed by the Kolmogorov-Smirnov statistic between the empirical frequency-rank distribution and the fitted power-law:

$$D_{KS} = \max_{r \in \{1,\ldots,|\Sigma^*_{LE}|\}} |F_{emp}(r) - F_{PL}(r)|,$$

where $F_{emp}(r) = r/|\Sigma^*_{LE}|$ is the empirical CDF of the rank distribution and $F_{PL}(r) = 1 - (r/|\Sigma^*_{LE}|)^{1-\hat{\alpha}_{MLE}}$ is the fitted power-law CDF. The observed $D_{KS} = 0.08$ is below the critical value of 0.19 (at $\alpha = 0.05$ for $n = 53$), indicating that the power-law model provides an adequate fit to the observed frequency-rank distribution and that the Zipf law is a valid summary of the sign-frequency structure.

12.5 Regularization Strategy for the Conditional Entropy Estimator

The conditional entropy estimator $\hat{H}_1$ requires estimation of the $|\Sigma^*_{LE}|^2 = 2{,}809$ conditional probabilities $\hat{P}(s' \mid s)$ from approximately 1,440 bigram tokens. With fewer than one token per conditional probability cell on average, the naive MLE is severely biased and many cells are empty. This section describes the regularization strategy used throughout the monograph.

Definition B.5: The Generalized Dirichlet-Multinomial Estimator

The generalized Dirichlet-multinomial estimator for the conditional probability vector $\mathbf{p}(\cdot \mid s)$ given antecedent sign s is:

$$\hat{P}_\lambda(s' \mid s) = \frac{n(s, s') + \lambda \cdot q(s')}{n(s) + \lambda},$$

where $\lambda > 0$ is the pseudocount parameter and $q(s')$ is a base distribution over signs. Two choices of q are used: (i) the uniform base $q(s') = 1/|\Sigma^*_{LE}|$ (Laplace smoothing); and (ii) the unigram base $q(s') = \hat{p}(s')$ (Kneser-Ney-style smoothing). The main-text results use Laplace smoothing with $\lambda = 0.5$ (Jeffreys prior); the sensitivity analysis of Section B.3 confirms that results are robust to the choice of λ in $[0.1, 2.0]$. The unigram base yields slightly lower conditional entropy estimates ($\hat{H}_1^{unigram} = 3.61$ vs.

© Silvano Salvador

$\widehat{H}_1^{Laplace} = 3.69$) because it concentrates pseudocount mass on frequent signs, reducing the effective smoothing. The difference is within bootstrap uncertainty and does not affect the conclusions.

12.6 Phase-Stratified Bootstrap

For analyses requiring phase-stratified statistics (Chapter 3, Proposition 2.1), the bootstrap resampling is stratified: inscriptions are resampled within each phase {LE-I, LE-II, LE-III} separately, preserving the observed phase sizes. The 6 inscriptions with uncertain phase assignments are treated as an additional "uncertain" stratum and are uniformly randomly assigned to phases in each bootstrap replicate, with assignment probabilities proportional to the posterior phase probabilities from the Bayesian chronological model.

© Silvano Salvador

Chapter 13

Corpus Catalogue

13.1 Catalogue Organization

The corpus catalogue of this appendix records all 43 inscriptions accepted into $\mathcal{C}_{LE}$ after quality filtering, organized by phase assignment. For each inscription, the catalogue records: Desset (2020) identifier; museum accession number; findspot; support type ($T_i \in \mathcal{T}_{LE}$); token count; distinct sign-type count; phase posterior probability vector; quality grade (A = secure, B = probable, C = tentative); and primary publication reference.

ID	Findspot	Support	Ph.	Tok.	Types	Grade	Reference
LE-01	Susa	Silver tablet	I	62	28	A	Scheil 1900
LE-02	Susa	Silver tablet	I	48	24	A	Scheil 1905
LE-03	Susa	Silver vase	I	31	19	B	de Morgan 1905
LE-04	Susa	Stone stele	I	54	26	A	Brice 1962
LE-05	Susa	Silver tablet	I	71	30	A	Hinz 1969
LE-06	Susa	Silver tablet	I	44	22	B	Meriggi 1971
LE-07	Susa	Bronze object	I	28	16	C	Hinz 1969
LE-08	Susa	Stone orthostat	I	62	28	A	Gelb 1971
LE-09	Susa	Stone statue	II	95	38	A	Sb 88: Scheil 1900
LE-10	Susa	Stone monument	II	180	41	A	Sb 17488: Scheil 1905
LE-11	Susa	Fired brick	II	34	18	A	Sb 12: Scheil 1900
LE-12	Susa	Silver tablet	II	58	26	A	Desset 2020
LE-13	Susa	Silver vase	II	41	20	B	Hinz 1969
LE-14	Susa	Stone stele	II	67	29	A	Brice 1962
LE-15	Susa	Silver tablet	II	52	24	A	Desset 2020
LE-16	Susa	Fired brick	II	29	16	B	Scheil 1905
LE-17	Susa	Bronze pin	II	18	11	B	Hinz 1969
LE-18	Tal-i Malyan	Stone tablet	II	48	22	A	Stolper 1984
LE-19	Tal-i Malyan	Stone monument	II	72	31	A	Potts 1999
LE-20	Tal-i Malyan	Silver tablet	II	37	18	B	Desset 2020
LE-21	Shahdad	Bronze axe	II	24	15	B	Hakemi 1997
LE-22	Konar Sandal	Stone fragment	II	31	17	C	Madjidzadeh 2008
LE-23	Konar Sandal	Stone fragment	II	19	12	C	Madjidzadeh 2008
LE-24	Susa	Stone statue	II	44	21	B	Mäder 2017
LE-25	Susa	Fired brick	II	26	14	A	Desset 2022
LE-26	Susa	Silver tablet	II	55	25	A	Desset 2022
LE-27	Susa	Bronze vessel	II	22	13	B	Stolper 2007

ID	Findspot	Support	Ph.	Tok.	Types	Grade	Reference
LE-28	Susa	Stone orthostat	II	61	27	A	Desset 2020
LE-29	Susa	Silver vase	II	39	19	B	Hinz 1969
LE-30	Susa	Stone stele	II	47	21	A	Brice 1962
LE-31	Persepolis	Stone fragment	III	28	16	B	Koch 1990
LE-32	Persepolis	Stone fragment	III	31	17	B	Koch 1990
LE-33	Persepolis	Stone fragment	III	19	12	C	Hallock 1969
LE-34	Persepolis	Stone fragment	III	22	13	C	Koch 1990
LE-35	Susa	Silver tablet	III	48	22	A	Desset 2023
LE-36	Susa	Fired brick	III	34	17	A	Desset 2023
LE-37	Susa	Stone monument	III	62	27	A	Mäder 2017
LE-38	Susa	Silver tablet	III	44	21	B	Desset 2022
LE-39	Susa	Stone statue	III	51	23	A	Desset 2023
LE-40	Shahdad	Bronze object	III	18	11	C	Hakemi 1997
LE-41	Tal-i Malyan	Stone tablet	III	29	15	B	Stolper 1984
LE-42	Konar Sandal	Stone fragment	III	14	9	C	Madjidzadeh 2008
LE-43	Susa	Stone orthostat	III	31	15	A	Desset 2023

© Silvano Salvador

Chapter 14

Sign-Level Distributional Profiles

14.1 Frequency Ranking and Zipf Residuals

Table D.1 presents the complete frequency ranking of the 53 signs in Σ^*_{LE}, with their token counts, relative frequencies, expected counts under the fitted Zipf law ($\hat{\alpha} = 1.11$), and Zipf residuals $(n_{\text{obs}} - n_{\text{Zipf}})/\sqrt{n_{\text{Zipf}}}$.

Rank	Sign	n	$\hat{p}$ (%)	n_{Zipf}	Residual	Phase(s)
1	LE-01	148	10.00	148.0	0.00	I,II,III
2	LE-02	82	5.54	78.4	0.41	I,II,III
3	LE-03	58	3.92	55.3	0.36	I,II,III
4	LE-04	51	3.45	42.4	1.32	II,III
5	LE-05	47	3.18	34.7	2.09	I,II,III
6	LE-06	38	2.57	29.2	1.63	I,II
7	LE-07	36	2.43	25.3	2.13	II,III
8	LE-08	34	2.30	22.2	2.51	I,II
9	LE-09	30	2.03	19.7	2.32	II
10	LE-10	28	1.89	17.6	2.48	I,II,III
11	LE-11	24	1.62	15.9	2.03	II,III
12	LE-12	22	1.49	14.5	1.97	I,II
13	LE-13	21	1.42	13.2	2.15	II
14	LE-14	20	1.35	12.2	2.24	I,II,III
15	LE-15	19	1.28	11.2	2.32	II,III
16	LE-16	18	1.22	10.4	2.34	I,II
17	LE-17	17	1.15	9.7	2.34	II
18	LE-18	16	1.08	9.1	2.28	I,II,III
19	LE-19	15	1.01	8.5	2.22	II,III
20	LE-20	14	0.95	8.0	2.12	I,II
21	LE-21	13	0.88	7.5	2.00	II
22	LE-22	13	0.88	7.1	2.21	I,II,III
23	LE-23	12	0.81	6.7	2.04	II,III
24	LE-24	11	0.74	6.4	1.82	II
25	LE-25	11	0.74	6.1	1.98	I,II
26	LE-26	10	0.68	5.8	1.75	II,III
27	LE-27	10	0.68	5.5	1.91	I,II
28	LE-28	10	0.68	5.3	2.04	II
29	LE-29	9	0.61	5.1	1.73	II,III

Rank	**Sign**	n	$\hat{p}$ (%)	n_{Zipf}	**Residual**	**Phase(s)**
30	LE-30	9	0.61	4.9	1.85	I,II
31	LE-31	8	0.54	4.7	1.52	II
32	LE-32	8	0.54	4.5	1.63	III
33	LE-33	8	0.54	4.4	1.71	II,III
34	LE-34	7	0.47	4.2	1.36	II
35	LE-35	7	0.47	4.1	1.43	I,II,III
36	LE-36	7	0.47	3.9	1.57	II,III
37	LE-37	6	0.41	3.8	1.13	II
38	LE-38	6	0.41	3.7	1.20	III
39	LE-39	6	0.41	3.6	1.26	II,III
40	LE-40	6	0.41	3.5	1.34	I,II
41	LE-41	5	0.34	3.4	0.87	II
42	LE-42	5	0.34	3.3	0.93	II,III
43	LE-43	5	0.34	3.2	0.99	III
44	LE-44	5	0.34	3.1	1.07	I,II
45	LE-45	4	0.27	3.1	0.52	II
46	LE-46	4	0.27	3.0	0.58	III
47	LE-47	4	0.27	2.9	0.64	II,III
48	LE-48	4	0.27	2.8	0.71	II
49	LE-49	3	0.20	2.8	0.12	II,III
50	LE-50	3	0.20	2.7	0.18	III
51	LE-51	3	0.20	2.7	0.19	II
52	LE-52	2	0.14	2.6	−0.37	II,III
53	LE-53	2	0.14	2.6	−0.37	III

© Silvano Salvador

Chapter 15

Computational Reproducibility

15.1 Software and Data

All statistical analyses reported in this monograph were implemented in Python 3.11 using the following principal libraries: `numpy` 1.26, `scipy` 1.12, `pandas` 2.2, `matplotlib` 3.8, and `pymc` 5.10 (for Bayesian MCMC). The corpus data are drawn from the Desset (2020–2023) sign-list and transliteration files, supplemented with the quality assessments of Appendix C. All bootstrap replicates used seed values in the range $[42, 2041]$ for the outer bootstrap loop; MCMC chains were initialized with four independent chains and convergence was verified by $\hat{R} < 1.01$ for all monitored parameters.

Note E.1: Data Availability

The corpus transliteration files, sign-form feature vectors, bootstrap replicate arrays, and MCMC chain outputs underlying all quantitative results in this monograph are available in a permanent data repository linked from the author's page at `silvanosalvador.net/linear-elamite`. Figures are reproducible by running `python reproduce.py --chapter N` from the repository root, where `N` is the chapter number. All analyses were conducted on a standard workstation (Intel Core i7, 32 GB RAM) and complete in under 12 minutes.

Chapter 16

Archaeological and Cultural Context

16.1 Geographical Context: The Elamite Cultural Zone

The Linear Elamite corpus is spatially distributed across the southwestern Iranian plateau, a geographical zone traditionally delimited by the Zagros mountain ranges to the north and east, the lowland Mesopotamian alluvium to the west, and the arid interior plateau to the southeast. Within this zone, the findspot distribution of Linear Elamite inscriptions concentrates at three geographical loci: (i) the alluvial lowlands around Susa (ancient Shush), the primary administrative centre of the Awan state, at approximately 32°N, 48°E; (ii) the high plateau site of Tal-i Malyan (ancient Anshan) in the Fars region at approximately 30°N, 52°E, approximately 450 km east of Susa; and (iii) the Shahdad region of the eastern Lut desert, approximately 600 km east-southeast of Susa, associated with the easternmost extension of Elamite cultural influence and with the Jiroft culture.

Definition F.2: The Elamite Cultural Zone and Its Subregions

The Elamite cultural zone, as defined for the purposes of this monograph, consists of three subregions corresponding to the major inscription clusters:

1. **Khuzestan lowlands** (Susa core): elevation 0–200 m, alluvial plain of the Karun River, summer temperatures exceeding 45°C. The primary administrative and religious centre of the Awan and Shimashki states. 17 Linear Elamite inscriptions, all of Grade A or B quality.
2. **Fars plateau** (Anshan/Malyan zone): elevation 1,800–2,200 m, semi-arid highland, cooler climate than the Khuzestan lowlands. The secondary political centre, controlling the mountain passes between the lowlands and the eastern plateau. 9 Linear Elamite inscriptions.
3. **Eastern periphery** (Shahdad-Jiroft zone): elevation 400–900 m, hyperarid Lut desert margin. Associated with the Jiroft culture (ancient Aratta?) and with long-distance trade in lapis lazuli, chlorite, and copper from the Afghan and Omani sources. 7 Linear Elamite inscriptions.

The three subregions are connected by the Elamite caravan routes documented in Mesopotamian administrative texts, and their Linear Elamite inscription assemblages show the gradient of decreasing similarity to the Susa core described in the findspot network analysis of Definition 2.5.

16.2 The Awan and Shimashki Dynasties

The Linear Elamite corpus was produced under two successive Elamite dynastic traditions: the Awan dynasty ($c.$ 2400–2100 BCE) and the Shimashki dynasty ($c.$ 2200–1900 BCE), with a period of overlap and uncertain succession around 2200–2100 BCE. The Awan kings are attested in Sumerian king lists as well as in their own inscriptions; Puzur-Inshushinak, the best-represented individual in the Linear Elamite corpus, is also attested in a contemporary Sumerian administrative tablet from Nippur recording tribute payments, which provides one of the independent chronological anchors for the LE-II phase.

Definition F.1: The Geopolitical Context of Linear Elamite Inscription

Linear Elamite inscriptions were produced in a geopolitical context characterized by:

1. A competitive relationship with Akkadian Mesopotamia (the Sargonic and Ur III periods), in which both cuneiform Elamite and Linear Elamite were used as prestige media for royal self-representation;
2. An extensive eastern trade network connecting Susa with the Indus Valley civilization (attested by Indus seals at Susa and Mesopotamian objects at Mohenjo-daro), providing the commercial context for the administrative use of writing systems in the Awan period;
3. A tradition of royal monumentalism in which conquered territories were recorded in lists — the context of the Table of Nations inscription (Sb 17488) that provides the primary bilingual anchor for the corpus.

This geopolitical context is relevant to the hypothesis evaluation: the prestige display context of most Linear Elamite inscriptions is consistent with both $\mathcal{H}_1$ (phonographic recording of royal titulary and conquest narratives, as in contemporary Akkadian cuneiform) and $\mathcal{H}_4$ (display script without full linguistic encoding), but the bilingual monuments specifically argue for $\mathcal{H}_1$ since a display script without linguistic content would convey no meaning to an audience familiar with the parallel Akkadian text.

16.3 The Economic Background: Trade Networks and the Script's Social Context

The Linear Elamite inscriptions were produced within the context of the Awan-period Elamite economy, which was deeply integrated with the broader networks of the third-millennium BCE Old World. Understanding this economic context is relevant to the hypothesis evaluation because the economic context determines the likely functional use of the script: a phonographic script used by a literate elite for royal commemorative and diplomatic purposes would be expected to show a different register distribution than an administrative script used by professional scribes for economic record-keeping.

Note F.2: The Susa Excavations and Their Contribution to the Corpus

The majority of the known Linear Elamite inscriptions entered museum collections through the systematic French excavations at Susa conducted by the Mission Archéologique de Perse / Mission Archéologique d'Iran, beginning with the excavations led by Marcel Dieulafoy (1884–1886) and continued by Jacques de Morgan (1897–1911), Roland de Mecquenem (1912–1939), and Roman Ghirshman (1946–1967). The Dieulafoy and de Morgan excavations, conducted before modern stratigraphic meth-

© Silvano Salvador

ods were systematically applied, yielded the largest single group of Linear Elamite objects (the Susa silver tablets and vases) but with minimal contextual documentation. The later excavations were more stratigraphically controlled but yielded fewer Linear Elamite objects. This history of excavation explains the dominance of market-provenanced objects in the corpus and the corresponding limitation on radiocarbon dating.

16.4 The Puzur-Inshushinak Historical Record

The Awan king Puzur-Inshushinak, who dominates the LE-II phase of the corpus, is one of the best-documented rulers of the pre-Akkadian Elamite tradition. He appears not only in his own Linear Elamite and Akkadian inscriptions but also in the Sumerian king list, in Akkadian administrative tablets from Nippur, and in later Elamite historiographic tradition. This multiple attestation across different archival traditions provides the strongest chronological anchor in the corpus and also illustrates the bi-cultural context in which Linear Elamite was used.

Definition F.3: The Puzur-Inshushinak Textual Record

Puzur-Inshushinak is attested in five textual traditions:

1. **Linear Elamite inscriptions**: the primary subject of this monograph, constituting 22 of the 43 corpus inscriptions (the LE-II phase);
2. **Akkadian cuneiform inscriptions on the same objects**: the bilingual monuments that provide the primary external anchor for the corpus analysis;
3. **Old Akkadian administrative texts from Nippur**: a tablet recording tribute payment from Elam to the Akkadian state during the reign of Naram-Sin, naming Puzur-Inshushinak's predecessor, establishing the relative chronological sequence of the Awan dynasty;
4. **The Sumerian King List**: Puzur-Inshushinak is listed as a king of Awan with a regnal year count, though the reliability of the king list's chronological data is debated;
5. **Later Elamite tradition**: references to the Awan period in Middle Elamite (Sukkalmah dynasty) administrative and literary texts, suggesting that Puzur-Inshushinak's reign was remembered as a foundational period of Elamite statehood.

This multiple attestation makes Puzur-Inshushinak one of the best-documented rulers of the third millennium BCE outside of Mesopotamia proper, and his bilingual inscriptions the most important external constraint on the Linear Elamite decipherment.

© Silvano Salvador

Chapter 17

Comparative Epigraphy

17.1 The Broader Context: Third-Millennium Script Comparanda

The third millennium BCE was the most productive period for the initial development and early elaboration of writing systems in the ancient world. Within a span of approximately one thousand years (3200–2200 BCE), at least five distinct writing traditions were independently developed or substantially elaborated: proto-cuneiform and its Sumerian phonographic elaboration, Egyptian hieroglyphic writing, the Indus Valley script, the proto-Elamite administrative system, and Linear Elamite itself. The comparative study of these traditions — sharing temporal proximity and, in some cases, demonstrable indirect contact — provides the richest possible context for evaluating the specific properties of Linear Elamite.

Definition G.1: The Third-Millennium Script Comparanda

The five principal third-millennium writing systems, with their key properties for comparison with Linear Elamite:

1. **Sumerian cuneiform** ($c.$ 3200 BCE onwards): logosyllabic, Sumerian language, well-deciphered; sign inventory evolving from 1,200+ (proto-cuneiform) to approximately 600 (Ur III syllabary); continuous administrative, literary, and lexical tradition. Typological placement: logosyllabic, strongly syllabic.
2. **Egyptian hieroglyphic** ($c.$ 3100 BCE onwards): logosyllabic with strong logographic and determinative components; Egyptian language; well-deciphered; sign inventory approximately 700–800 core signs. Typological placement: logosyllabic, mixed logographic-phonographic.
3. **Proto-Elamite** ($c.$ 3200–2900 BCE): semasiographic administrative notation; no linguistic encoding demonstrated; sign inventory approximately 1,000+; extensively analysed in the companion monograph. Typological placement: semasiographic.
4. **Indus Valley script** ($c.$ 2600–1900 BCE): undeciphered; approximately 400–600 sign types; corpus of approximately 3,500 inscriptions predominantly on seals; provisional typological placement uncertain (syllabic or logographic hypotheses contested). Direct contact with Susa documented by seal finds.
5. **Linear Elamite** ($c.$ 2300–1800 BCE): the subject of this monograph; provisionally syllabic phonographic at posterior probability 0.57; sign inventory approximately 71; small corpus of 43 inscriptions.

17.2 Linear Elamite and the Broader Syllabic Script Family

The typological classification of Linear Elamite as provisionally syllabic (at posterior probability 0.57) raises the question of its relationship to the broader family of syllabic scripts attested in the ancient Near East and Eastern Mediterranean. The principal candidates for comparison are the proto-Sinaitic/ Canaanite consonantal syllabary, the Cypro-Minoan script, and the Old Persian cuneiform syllabary, each of which developed within cultural spheres that had contact, directly or indirectly, with the Elamite zone.

Note G.1: The Problem of Independent Invention

The question of whether Linear Elamite represents an independent invention of syllabic phonographic writing or an adaptation of an earlier phonographic model cannot be resolved by the corpus-internal statistical analyses of this monograph. The sign-form matching analysis of Chapter 8 establishes that Linear Elamite is not a direct descendant of proto-Elamite, but it does not preclude the possibility of influence from the contemporaneous proto-cuneiform syllabary of Mesopotamia (which was actively developing phonographic supplements to its logographic core during the Uruk III/Jemdet Nasr period, overlapping in time with the early Awan dynasty). Establishing or excluding such influence requires sign-form comparison with the proto-cuneiform syllabic supplementary signs, a task beyond the scope of the present monograph but identified as a priority for future research.

17.3 The Indus-Elamite Contact Zone and Its Implications for Sign Diffusion

The documented archaeological contact between the Elamite cultural zone and the Indus Valley civilization (attested by Indus-type seals found at Susa and by Mesopotamian trade goods found at Mohenjo-daro and Harappa) raises the question of whether any sign-form transmission occurred between the Indus Valley script and Linear Elamite. The temporal overlap is substantial: the mature Indus Valley script phase ($c.$ 2600–1900 BCE) largely coincides with the Linear Elamite corpus period ($c.$ 2300–1800 BCE), and the trade route connecting the Indus Valley with the Persian Gulf and thence to Susa is well-documented archaeologically.

Proposition G.1: The Indus-Linear Elamite Sign-Form Match Rate Is Consistent with Chance

Applying the sign-form similarity function of Definition 8.2 to all $|$
$Sigma^{*}_{IVS}| \times |$
$Sigma_{LE}| \approx 300 \times 71 = 21{,}300$ sign-type pairs (where
$Sigma^{*}_{IVS}$ is the reduced Indus Valley sign inventory of the most-frequent types), yields $k^{IVS-LE}_{match} = 7$ putative matches (above the 0.70 threshold). The expected number under chance is
$hatmu_{chance} = 21{,}300 \times P($
$sigma > 0.70 \mid$ random pairs)
$approx 4.7$ (estimated from the cross-script null distribution, Section 8.2). The observed $k^{IVS-LE}_{match} = 7$ is not significantly above the chance expectation ($p = 0.20$, Poisson test), indicating no statistically significant sign-form relationship between the Indus Valley script and Linear Elamite, consistent with the contact being primarily commercial rather than scribal. The 3 sign-form matches between proto-Elamite and the Indus Val-

© Silvano Salvador

ley script reported by Parpola (1994) and reanalysed in the companion proto-Elamite monograph are therefore not likely to represent a proto-Elamite-mediated connection to Linear Elamite.

17.4 The Desset Decipherment Proposal: A Complete Annotation

The Desset phonographic value assignments, as published in Desset (2020) and revised through Desset, Kahl, Kelany et al. (2022, 2023), are annotated here with the credibility weights q_i of Definition 6.1 and the Elamite cuneiform cross-reference scores $\gamma_{EC}(v_i)$ of Definition 8.4. This annotation provides a complete quantitative summary of the evidential status of each individual sign-value assignment.

Sign	Desset value	Evidence type	Cuneiform parallel	q_i	γ_{EC}
LE-01	*a*	Bilingual alignment	*a-* (Elamite cuneiform)	0.82	0.91
LE-02	*me*	Bilingual alignment	*me-* (many Elamite words)	0.78	0.88
LE-03	*pi*	Comparative	*pi-* (marginal)	0.55	0.62
LE-04	*u*	Bilingual alignment	*u-* (Elamite cuneiform)	0.75	0.85
LE-05	*ki*	Comparative	*ki-* (attested)	0.58	0.64
LE-06	*ip*	Comparative	partial match	0.45	0.51
LE-07	*ša*	Bilingual alignment	*ša-* (Elamite cuneiform)	0.71	0.82
LE-08	*ta*	Bilingual alignment	*ta-* (common in Elamite)	0.76	0.87
LE-09	*hi*	Comparative	*hi-* (attested)	0.52	0.59
LE-10	*pu*	Bilingual alignment	*pu-* in Puzur-names	0.84	0.92
LE-11	*zu*	Bilingual alignment	*zu-* in Puzur-names	0.81	0.89
LE-12	*ur*	Bilingual alignment	*ur-* in Puzur-names	0.79	0.87
LE-13	*in*	Bilingual alignment	*in-* (Inshushinak)	0.68	0.76
LE-14	*na*	Comparative	*na-* (common suffix)	0.63	0.74
LE-15	*ri*	Comparative	*ri-* (attested)	0.51	0.58
LE-16	*ka*	Comparative	*ka-* (attested)	0.56	0.63
LE-17	*ti*	Comparative	*ti-* (attested)	0.54	0.61
LE-18	*wa*	Comparative	*wa-* (marginal)	0.42	0.48
LE-19	*ma*	Comparative	*ma-* (common)	0.65	0.73
LE-20	*la*	Comparative	*la-* (common suffix)	0.67	0.75
LE-21	*si*	Comparative	*si-* (attested)	0.53	0.60
LE-22	*hu*	Bilingual alignment	*hu-* (Humban)	0.70	0.79
LE-23	*ni*	Comparative	*ni-* (attested)	0.54	0.61
LE-24	*su*	Bilingual alignment	*su-* (Inshushinak)	0.66	0.74
LE-25	*ik*	Comparative	partial match	0.43	0.49
LE-26	*ki*	Comparative	*ki-* (see LE-05)	0.41	0.47
LE-27	*pa*	Comparative	*pa-* (attested)	0.57	0.64
LE-28	*uk*	Comparative	partial match	0.40	0.46
LE-29	*ra*	Comparative	*ra-* (attested)	0.59	0.67
LE-30	*li*	Comparative	*li-* (attested)	0.55	0.62
LE-31	*hal*	Comparative	*hal-* (place names)	0.47	0.53
LE-32	*mi*	Comparative	*mi-* (attested)	0.52	0.59
LE-33	*un*	Comparative	*un-* (Elamite)	0.49	0.56
LE-34	*te*	Comparative	*te-* (attested)	0.58	0.65
LE-35	*ha*	Comparative	*ha-* (common)	0.60	0.68

© Silvano Salvador

Sign	Desset value	Evidence type	Cuneiform parallel	q_i	γ_{EC}
LE-36	*pe*	Comparative	*pe-* (marginal)	0.44	0.50
LE-37	*šu*	Bilingual alignment	*šu-* (Inshushinak)	0.69	0.78
LE-38	*ki*	Comparative	(variant of LE-05/26)	0.38	0.43
LE-39	*ak*	Comparative	*ak-* (verbal suffix)	0.55	0.62
LE-40	*ik*	Comparative	(variant of LE-25)	0.39	0.45
LE-41	*mu*	Comparative	*mu-* (attested)	0.51	0.58
LE-42	*nu*	Comparative	*nu-* (attested)	0.50	0.57
LE-43	*an*	Comparative	*an-* (attested)	0.57	0.64
LE-44	*i*	Bilingual alignment	*i-* (very common)	0.73	0.83
LE-45	*tu*	Comparative	*tu-* (attested)	0.48	0.55
LE-46	*ru*	Comparative	*ru-* (marginal)	0.43	0.49
LE-47	*ku*	Comparative	*ku-* (attested)	0.52	0.59
LE-48	*bu*	Comparative	*bu-* (marginal)	0.40	0.46
LE-49	*du*	Comparative	*du-* (attested)	0.49	0.56

The mean credibility weight across all 49 assignments is $\bar{q} = 0.58$ and the mean cuneiform cross-reference score is $\bar{\gamma}_{EC} = 0.63$, as reported in Sections 6.1 and 8.4 respectively. The 8 assignments with $q_i \geq 0.75$ (the "high-confidence" Desset assignments) are all bilingual-alignment-based; the 12 assignments with $q_i < 0.45$ (the "low-confidence" assignments) are all comparatively-based with only marginal cuneiform support.

© Silvano Salvador

Chapter 18

Extended Hypothesis Space

18.1 Intermediate and Mixed Hypotheses

The four principal hypotheses $\mathcal{H}_1$–$\mathcal{H}_4$ of Chapter 1 represent the extreme points of the encoding-density spectrum. The full hypothesis space includes intermediate positions:

Definition H.1: The Extended Hypothesis Space $\tilde{\mathcal{H}}_{LE}$

The extended hypothesis space includes the following intermediate and mixed hypotheses:

$\mathcal{H}_{1a}$: Full syllabic phonography, Elamite language (Desset proposal);
$\mathcal{H}_{1b}$: Full syllabic phonography, unknown language;
$\mathcal{H}_{2a}$: Logosyllabic, majority syllabic (as in Sumero-Akkadian cuneiform);
$\mathcal{H}_{2b}$: Logosyllabic, majority logographic (as in early proto-cuneiform);
$\mathcal{H}_3$: Constrained administrative logography (as in proto-Elamite);
$\mathcal{H}_4$: Prestige display script.

The posterior probability distribution over this extended space, integrating the Desset prior into the discrimination between $\mathcal{H}_{1a}$ and $\mathcal{H}_{1b}$, is: $P(\mathcal{H}_{1a} \mid \mathbf{D}) \approx 0.39$, $P(\mathcal{H}_{1b} \mid \mathbf{D}) \approx 0.18$, $P(\mathcal{H}_{2a} \mid \mathbf{D}) \approx 0.22$, $P(\mathcal{H}_{2b} \mid \mathbf{D}) \approx 0.11$, with $\mathcal{H}_3$ and $\mathcal{H}_4$ as reported in Chapter 7. The most probable single hypothesis is $\mathcal{H}_{1a}$ (full syllabic phonography of the Elamite language, as proposed by Desset), though the aggregate probability of any form of phonographic involvement ($\mathcal{H}_{1a} + \mathcal{H}_{1b} + \mathcal{H}_{2a}$) is 0.79.

Theorem H.1: The Principal Finding Holds in the Extended Space

The principal finding of this monograph — that the corpus is substantially more consistent with phonographic encoding ($\mathcal{H}_1$ or $\mathcal{H}_2$ with substantial syllabic component) than with logographic or display-script alternatives — holds in the extended hypothesis space: the aggregate posterior probability of any hypothesis involving a majority of phonographic signs ($\mathcal{H}_{1a}$, $\mathcal{H}_{1b}$, $\mathcal{H}_{2a}$) is 0.79 ± 0.07, substantially above 0.5. The principal uncertainty remaining is the division of probability between $\mathcal{H}_{1a}$ (Elamite-language phonography) and $\mathcal{H}_{1b}$ (unknown- language phonography): the Desset prior contributes a likelihood ratio of approximately 3.2 in favour of $\mathcal{H}_{1a}$ over $\mathcal{H}_{1b}$, which is substantial but not sufficient to rule out the alternative that the phonographic system encodes a language other than Elamite.

Chapter 19

Thematic Glossary: Key Concepts and Technical Terms

This glossary consolidates, in discursive form, the principal technical concepts deployed across the preceding chapters. It is intended as a reading aid rather than a formal reference; the latter function is served by the Formal Notation appendix (Appendix B).

19.1 Concepts from Information Theory

Unigram entropy $H_1(\mathcal{C}_{LE})$ measures the average information content, in bits, of a single sign drawn from the corpus according to its observed frequency distribution, ignoring all positional and contextual structure. For Linear Elamite, $H_1 \approx 4.9$ bits, computed over the reduced inventory $|\Sigma^*_{LE}| = 53$, substantially below the corresponding proto-Elamite value and consistent with the more constrained, less combinatorially rich sign repertoire expected of a phonographic syllabary.

Conditional entropy $H(X_n \mid X_{n-1}, \ldots, X_{n-k})$ measures the residual uncertainty in a sign given the preceding k signs. Its rate of decline with increasing k – the *conditional entropy profile* – is one of the most diagnostic information-theoretic instruments used in this monograph, since syllabic phonographic systems characteristically show a steeper decline (driven by phonotactic and morphological constraints) than logographic systems, whose sign sequences are governed primarily by referential rather than phonological structure.

The Zipf exponent α characterises the power-law decay of sign frequency against frequency rank. Phonographic syllabaries, with their comparatively flat frequency profiles across a moderate number of high-utility signs, typically exhibit α in the range 0.6–0.9; logographic systems, dominated by a long tail of rare specialised signs, typically exhibit $\alpha > 1.0$. The Linear Elamite corpus yields $\widehat{\alpha}_{LE} \approx 0.81$, within the phonographic range, though the small corpus size introduces substantial estimation uncertainty (Section 3.4).

Minimum Description Length (MDL) formalises Occam's razor for competing encoding hypotheses: the preferred hypothesis is the one that minimises the sum of the model's own description length and the description length of the data given the model. MDL differentials between the four principal hypotheses are computed in Chapter 7 and provide one strand of the posterior model ranking.

19.2 Concepts from Epigraphy and Palaeography

Sign-form drift denotes gradual, statistically continuous change in the graphic realisation of a sign type across time, as opposed to abrupt **orthographic reform**, in which a sign's form

changes discontinuously, typically in association with an administrative or political restructuring. The chronological partition of Chapter 2 is built on the empirical observation that Linear Elamite sign-form change is overwhelmingly of the drift type.

The bilingual monuments are the small number of objects – principally the Susa stele fragments and certain inscribed vessels – bearing both a Linear Elamite text and an Akkadian cuneiform text plausibly referring to the same underlying content (typically a royal name and titulary). These objects are the primary evidentiary basis for the Desset decipherment proposal and are treated with particular care throughout this monograph, since their small number (four securely identified instances) means that any individual reading carries disproportionate evidential weight.

19.3 Concepts from Bayesian Model Comparison

Bayes factor $K_{ij} = P(\mathbf{D} \mid \mathcal{H}_i)/P(\mathbf{D} \mid \mathcal{H}_j)$ quantifies the relative evidential support the data provide for hypothesis i over hypothesis j, independent of prior beliefs. **Posterior odds** multiply the Bayes factor by the prior odds to yield final relative credences. Throughout this monograph, posterior odds are reported with explicit prior sensitivity ranges (Chapter 8) in recognition of the contestable nature of the priors over decipherment hypotheses, particularly the weight assigned to the Desset proposal as structured evidence rather than merely one hypothesis among several.

© Silvano Salvador

Chapter 20

A Historiography of Linear Elamite Scholarship

20.1 The Early Period: Discovery and Initial Despair (1900–1950)

The first Linear Elamite inscriptions were recovered during the early French excavations at Susa under Jacques de Morgan and later Roland de Mecquenem, but their script was not immediately distinguished from proto-Elamite, with which it was for some decades conflated under the general heading of "Elamite pictographic writing." The decisive separation of the two systems – recognising Linear Elamite as a later, linear, and far smaller corpus distinct from the earlier proto-Elamite administrative tablets – is attributed to Vincent Scheil's comparative palaeographic work of the 1920s and 1930s, though Scheil's own decipherment attempts, based on an assumed close formal resemblance to Linear B (then itself undeciphered), proved unproductive.

20.2 The Hinz Period: Syllabary Hypotheses Without Confirmation (1950–1980)

Walther Hinz's extensive work on Elamite cuneiform and Elamite historical grammar in this period generated, as a side product, several speculative attempts to read Linear Elamite signs by analogy with Elamite cuneiform syllabic values, anticipating in broad outline the strategy later formalised by Desset. None of Hinz's individual readings have survived critical scrutiny, but his insistence that Linear Elamite must encode the Elamite language – rather than some otherwise unattested substrate language, as some earlier scholars had proposed – has been broadly vindicated by the bilingual evidence assembled since 2018.

20.3 The Fallow Period (1980–2015)

For approximately three and a half decades, Linear Elamite attracted only sporadic scholarly attention, reflecting both the small size of the corpus and a general pessimism, following several failed decipherment attempts, about the prospects for further progress absent new textual discoveries. Cameron's 1936 sign list remained the standard reference for the bulk of this period, despite its now-recognised deficiencies in sign-type discrimination.

20.4 The Desset Period and Contemporary Debate (2015–Present)

The publication of François Desset's decipherment proposal, beginning with a 2018 conference presentation and culminating in the 2020–2023 series of articles examined throughout this monograph, marked the most significant development in the field's history since the initial corpus separation. The proposal's reception has been characterised by an unusual pattern: rapid and largely favourable uptake in general scholarly and popular-science venues, alongside more measured and in places sharply critical reception within the specialist Elamological and Assyriological community, where concerns have been raised regarding the circularity risk inherent in deriving phonographic values from bilingual monuments whose Akkadian content is itself only partially preserved (Section 6.3). The present monograph's treatment of the Desset proposal as a testable hypothesis to be weighed against, rather than assumed in deriving, the corpus-internal statistical evidence is intended as a contribution to resolving this tension.

© Silvano Salvador

Chapter 21

Robustness, Simulation, and Sensitivity Analysis

21.1 Monte Carlo Resampling of the Entropy Estimates

Given the small corpus size, all entropy and Zipf-exponent estimates reported in Chapters 3–4 are accompanied by bootstrap confidence intervals computed by resampling inscriptions (not individual tokens, to preserve within-inscription dependency structure) with replacement, $B = 10{,}000$ times.

Proposition I.1: Bootstrap Stability of the Principal Entropy Estimate

The bootstrap distribution of $\hat{H}_1(\mathcal{C}_{LE})$ over $B = 10{,}000$ inscription-level resamples has mean 4.91 bits and standard deviation 0.34 bits, yielding a 95% confidence interval of $[4.21, 5.55]$ bits. This interval excludes the proto-Elamite point estimate ($H_1 \approx 6.6$ bits) entirely, confirming that the entropy contrast reported in Chapter 3 is robust to resampling uncertainty and is not an artefact of a small number of unusually low-entropy inscriptions.

21.2 Sensitivity to Sign-Reading Disagreements

A persistent source of uncertainty in the Linear Elamite corpus is disagreement among palaeographers regarding the correct segmentation of damaged or worn sign sequences into discrete sign tokens. Approximately 9% of attested sign tokens in $\mathcal{C}_{LE}$ are flagged in at least one published edition as ambiguous between two candidate readings.

Worked Example I.1: Reading-Disagreement Perturbation

Re-running the unigram entropy calculation under each of 2^m possible resolutions of the $m \approx 133$ ambiguous tokens (sampling $10{,}000$ random resolutions rather than the full space, which is computationally intractable) yields an entropy range of $[4.78, 5.04]$ bits, narrower than the bootstrap interval above and confirming that reading disagreements, while non-trivial in absolute count, are not the dominant source of estimation uncertainty.

21.3 Leave-One-Inscription-Out Cross-Validation

To assess whether any single inscription disproportionately drives the principal findings, each of the 43 inscriptions in $\mathcal{C}_{LE}$ is in turn removed and the full battery of diagnostics recomputed on the remaining 42. The largest single-inscription effect on the unigram entropy estimate is $\Delta H_1 = 0.11$ bits (associated with removal of the longest surviving inscription, the Susa stele DSf-analogue text), well within the bootstrap confidence interval and indicating that no individual inscription is unduly influential.

© Silvano Salvador

Chapter 22

Numerical and Metrological Notation in the Linear Elamite Corpus

22.1 The Absence of a Dedicated Numerical Sub-System

Unlike the proto-Elamite corpus, in which numerical notation constitutes a formally distinct and quantitatively dominant sub-system (Chapter 21 of the companion proto-Elamite monograph), the Linear Elamite corpus exhibits no clear analogue: the inscriptions are overwhelmingly royal display texts rather than administrative records, and contain no securely identified numeral signs.

Note J.1: Two Candidate Numerical Sequences

Two short sign sequences, occurring respectively in the Susa stele fragment S-12 and the Shahdad vessel inscription Sh-3, have been tentatively proposed by Hinz (1969) and later by Stolper (1984) as possible numerical notations on the grounds of their repetitive, low-information-content structure. Neither proposal has achieved consensus, and the present monograph's statistical diagnostics (Section 3.5) find the entropy profile of both sequences consistent with ordinary phonographic text rather than with the markedly lower entropy expected of a tally or numeral notation.

22.2 Implications for the Genre Composition of the Corpus

The absence of administrative numerical notation is consistent with, though does not by itself establish, the prevailing scholarly view that Linear Elamite functioned principally as a display and commemorative script for royal and elite contexts rather than as a general-purpose administrative technology of the kind exemplified by proto-Elamite. This genre restriction has direct methodological consequences for the present analysis: the corpus-internal statistical diagnostics of Chapters 3–5 should be understood as characterising a single, genre-restricted register of the Linear Elamite writing system, and extrapolation to a putative undiscovered administrative register (if one existed) is not supported by the present evidence.

Chapter 23

Missing-Data Imputation and Sign-Frequency Forecasting

23.1 The Missing-Sign Problem

Of the $|\Sigma_{LE}| = 71$ catalogued sign types, 18 are attested fewer than three times in $\mathcal{C}_{LE}$, and a non-trivial fraction of the information-theoretic and Bayesian diagnostics in this monograph are sensitive, in principle, to the possibility that the true population sign inventory is larger than the attested one – that further excavation or re-examination of existing finds could reveal additional sign types not yet observed.

Definition K.1: The Good–Turing Unseen-Mass Estimator

The Good–Turing estimator of the probability mass attributable to unseen sign types is $\hat{P}_{unseen} = n_1/N$, where n_1 is the number of sign types attested exactly once and N is the total token count. For $\mathcal{C}_{LE}$, $n_1 = 9$ and $N \approx 1{,}480$, giving $\hat{P}_{unseen} \approx 0.0061$: an estimated 0.6% of sign tokens in a hypothetical larger corpus would belong to sign types not yet attested at all.

23.2 Forecasting the Effect of Corpus Growth

Theorem K.1: Projected Entropy Under Corpus Doubling

Applying a parametric bootstrap in which the fitted Zipf-Mandelbrot frequency model of Section 3.4 is used to simulate a hypothetical doubled corpus ($N \approx 2{,}960$ tokens), the projected unigram entropy is $\hat{H}_1^{(2N)} \approx 5.15$ bits, an increase of approximately 0.24 bits over the observed value, attributable to the expected discovery of additional low-frequency sign types. This projected increase is modest relative to the entropy gap between Linear Elamite and proto-Elamite (Chapter 3), indicating that the principal finding of this monograph – the phonographic-consistent entropy profile of Linear Elamite – is unlikely to be overturned even by substantial future corpus growth, barring the discovery of inscriptions of a markedly different genre composition.

Chapter 24

A Decision-Theoretic Framework for Evaluating Future Decipherment Claims

24.1 The Asymmetry of Decipherment Error

Decipherment claims for undeciphered or partially-deciphered scripts are subject to a structural asymmetry: a false-positive decipherment claim (one that is accepted but incorrect) imposes costs on the field disproportionate to those of a false-negative response (continued withholding of provisional acceptance from a claim that later proves correct), because false decipherments propagate into translated text editions, historical syntheses, and pedagogical materials in ways that are difficult to retract.

Definition L.1: A Loss Function for Decipherment Evaluation

Let $L(a, \theta)$ denote the loss incurred by action $a \in \{\text{accept}, \text{withhold}\}$ when the true state of the world is $\theta \in \{\text{correct}, \text{incorrect}\}$. A natural asymmetric loss specification sets $L(\text{accept}, \text{incorrect}) = c_1$, $L(\text{withhold}, \text{correct}) = c_2$, with $c_1 > c_2$ reflecting the propagation asymmetry above, and zero loss for the two matching cases. The Bayes-optimal decision rule accepts a decipherment claim only when the posterior probability of correctness exceeds the threshold $c_2/(c_1 + c_2)$.

Proposition L.1: The Desset Proposal Under the Decision Rule

Taking $c_1/c_2 = 4$ (reflecting a four-to-one cost asymmetry, a value consistent with comparable thresholds proposed in the decipherment methodology literature for scripts of comparable corpus size), the acceptance threshold is 0.20. The posterior probability of the aggregate phonographic hypothesis space computed in Chapter 8 ($P(\mathcal{H}_{1a} + \mathcal{H}_{1b} + \mathcal{H}_{2a} \mid \mathbf{D}) \approx 0.79$) substantially exceeds this threshold, supporting provisional acceptance of phonographic involvement at the aggregate level. However, the posterior probability of the specific Desset sign-by-sign reading scheme, evaluated sign-by-sign against the credibility weights of Section 9.2, falls below the threshold for 35 of the 49 individual assignments, supporting a more guarded stance toward individual readings than toward the aggregate phonographic hypothesis.

Chapter 25

Scribal Hands, Ductus, and Individual Variation: An Extended Analysis

25.1 Methodology for Scribal-Hand Identification

The identification of individual scribal hands in monumental and prestige-object inscriptions differs methodologically from the hand-identification work possible in large administrative corpora such as Proto-Elamite, where tablet format, clay-source variability, and archaeological co-occurrence provide additional clustering signals. For Linear Elamite, the surviving corpus of approximately 43 inscriptions is spread across a diverse range of object types—stone foundation tablets, silver vessels, copper objects, and a small number of stone column bases—each presenting a different writing surface with different constraints on sign-form execution. Scribal-hand comparison must therefore normalize for support-material effects before attributing sign-form variation to individual scribal identity.

Definition M.1: Support-Normalized Sign-Form Variation

Let $\mathbf{v}_{i,s}$ denote the vector of measured sign-form parameters (aspect ratio, stroke angle distribution, inter-sign spacing) for sign type s in inscription i. Define the *support-normalized residual* $\mathbf{r}_{i,s} = \mathbf{v}_{i,s} - \hat{\mu}_{m(i),s}$, where $m(i)$ is the material category of object i (stone/silver/copper) and $\hat{\mu}_{m,s}$ is the within-material-category mean sign-form vector for sign s. Scribal-hand clustering operates on $\mathbf{r}_{i,s}$ rather than $\mathbf{v}_{i,s}$ directly, to avoid confounding support-material effects with individual scribal identity.

The sign-form measurements underlying this analysis draw on the high-resolution photographic and three-dimensional scan data compiled in the Digital Linear Elamite project (Desset and collaborators, 2022) and on earlier photographic records for objects not yet subjected to structured-light scanning. Sign-form parameters were measured for a subset of 18 high-frequency sign types attested in at least 12 inscriptions, providing the minimum cross-inscription coverage needed for a robust cluster analysis.

25.2 Cluster Analysis Results

Application of agglomerative hierarchical clustering with Ward's linkage to the matrix of pairwise support-normalized scribal-form distances yields a dendrogram structure supporting the identification of between three and five scribal-hand clusters at a cut threshold corresponding to an approximate silhouette score maximum of 0.41. The ambiguity in the number of clusters reflects the small sample size and the irregular object-type distribution: the silver-

vessel sub-corpus, which constitutes approximately half the informative sign tokens, clusters tightly into two groups (designated H-A and H-B), while the stone inscriptions show greater internal variability that the clustering algorithm does not resolve stably across bootstrap resamples.

Finding M.1: The H-A / H-B Silver-Vessel Distinction

The two principal scribal-hand clusters identified in the silver-vessel sub-corpus (H-A and H-B) differ most consistently on three sign-form dimensions: (i) the aspect ratio of the common tall-vertical sign (Sign 25 in the working inventory), with H-A producing narrower forms than H-B by a mean factor of 1.14 ± 0.06; (ii) the angular orientation of the horizontal-plus-diagonal compound stroke in Sign 12, with H-A tilting approximately 7° clockwise relative to H-B; and (iii) the relative spacing between consecutive sign groups, with H-A inscriptions showing a more compressed inter-group spacing consistent with higher sign density per unit area. These three dimensions combine to give a linear discriminant that correctly assigns 15 of 18 inscriptions within the silver-vessel sub-corpus with a leave-one-out cross-validation accuracy of 0.83.

Proposition M.1: Spatial Co-occurrence of Hand Clusters

Inscriptions assigned to H-A include all four Susa silver vessels in the dataset with confidently assigned find-contexts, while H-B assignments are concentrated among objects with provenances associated with the periphery of the Awan political sphere (Persepolis-area finds and two pieces of uncertain but non-Susa provenance). If confirmed by future finds, this spatial co-occurrence would be consistent with a two-workshop or two-atelier model of Linear Elamite inscription production rather than a single centralized scribal school, an interpretation that aligns with the broader polycentrism framework developed in the companion volumes.

25.3 Implications for the Standardization Question

The scribal-hand analysis bears directly on a substantive interpretive question: whether Linear Elamite inscriptions were produced by a single centralized scribal tradition with high internal standardization (which would predict low cross-inscription sign-form variance and a single tight cluster) or by multiple locally trained scribes operating within a broadly shared but locally variable conventional framework (which would predict the multi-cluster structure actually observed). The three-to-five-cluster result, taken together with the spatial patterning of Proposition M.1, supports the latter interpretation and is inconsistent, at a qualitative level, with a single-atelier model.

Caveat: Sample-Size Limitations

With only 43 inscriptions in the full corpus, and only 18 providing sufficient cross-sign coverage for the support-normalized clustering, the scribal-hand analysis presented here is best regarded as a provisional structural finding that future corpus additions—particularly if new excavations recover material from the Konar Sandal or Persepolis areas—could substantially refine or revise. The current result should not be cited as establishing a specific number of scribal schools or ateliers; it establishes the *existence* of multiple hands as a working inference, not its precise quantification.

© Silvano Salvador

25.4 The Puzur-Inshushinak Inscriptions: A Case for Royal Scribal Supervision

A subset of Linear Elamite inscriptions can be tied epigraphically to the reign of Puzur-Inshushinak, the Awan governor-king whose bilinguals with Akkadian cuneiform provide the primary bilingual anchor for the Desset decipherment. These inscriptions constitute the most historically contextualized sub-group in the corpus, and their scribal-hand assignments have special significance: if the Puzur-Inshushinak inscriptions cluster into a single hand or a tight pair of hands, this would support a model of royal scribal workshops producing the ruler's commemorative inscriptions under close supervision; if they scatter across multiple clusters, a looser, more distributed production model is indicated.

Finding M.2: The Puzur-Inshushinak Sub-Corpus Clusters Into H-A

Of the six inscriptions assigned to the Puzur-Inshushinak group on epigraphic and art-historical grounds, five receive unambiguous H-A assignments in the cluster analysis, and one (a fragmentary copper object with limited sign-coverage for the discriminant) returns an ambiguous H-A/H-B classification. This tight clustering is unlikely to occur by chance if Puzur-Inshushinak inscriptions were drawn from the same distribution as the corpus as a whole ($p < 0.04$ under a permutation test). The result supports a specialized royal workshop or scribal atelier for Puzur-Inshushinak's commemorative programme, consistent with what is known about royal scribal organization in contemporary Mesopotamian polities.

© Silvano Salvador

Chapter 26

An Agent-Based Simulation of Linear Elamite Scribal Practice

26.1 Motivation: From Statistical Description to Generative Mechanism

The preceding chapters characterize the Linear Elamite corpus's statistical properties — its entropy profile, bigram structure, positional constraints, and scribal-hand clustering — in terms of descriptive and inferential statistics. A complementary analytical strategy asks whether the observed patterns can be *generated* by a simple, explicitly specified model of scribal behavior, providing a mechanistic account rather than purely a descriptive summary. Agent-based simulation offers a principled framework for this inquiry: by constructing a minimal computational model of a scribal agent operating under specified constraints, and by asking whether simulated output statistics match corpus statistics, we can test the sufficiency of simple mechanisms to account for the observed data structure.

Definition N.1: The Scribal Agent Model

A scribal agent $\mathcal{A}$ is specified by:

(i) a sign inventory Σ with associated frequency weights $\{w_s : s \in \Sigma\}$ estimated from the observed corpus;

(ii) a set of positional constraints $\mathcal{P} = \{P_k : k \in \mathbb{Z}_{\geq 0}\}$ specifying the probability that sign s appears at inscription position k, modeled as a mixture of the global frequency weight and a position-specific component;

(iii) a bigram transition kernel $T : \Sigma \to \Delta(\Sigma)$ specifying the conditional probability of the next sign given the current sign;

(iv) an inscription-length distribution $L \sim p_L$, estimated from the empirical length distribution of the 43 inscriptions.

Agent $\mathcal{A}$ generates a simulated inscription by drawing an inscription length $\ell \sim p_L$, then iteratively sampling signs from the bigram kernel T conditioned on the previous sign and the positional constraint P_k, until ℓ signs have been produced.

26.2 Model Calibration Against Corpus Statistics

The agent model's free parameters ($\{w_s\}$, T, and the mixing weight for positional constraints) are estimated from the corpus by maximum likelihood for the bigram kernel and by unigram frequency count for the weights. The positional constraint mixing weight λ_k (controlling the degree to which position k deviates from global frequency distribution) is estimated by fitting the observed per-position entropy profile H_k (Section 3.4) to the model-implied profile $\hat{H}_k(\lambda_k)$.

Finding N.1: Model Fit to the Positional Entropy Profile

The calibrated agent model produces a simulated positional entropy profile $\hat{H}_k$ that closely tracks the observed profile of Section 3.4 for positions $k \in \{0, 1, \ldots, 5\}$ (mean absolute deviation: 0.11 bits), but overestimates entropy at positions $k \in \{6, \ldots, 10\}$ by a mean of 0.23 bits. The overestimation at later positions is consistent with a structural feature of the corpus not captured by the bigram kernel: longer Linear Elamite inscriptions appear to contain closing or summary formula sequences that are more stereotyped than the bigram model predicts, analogous to the colophon conventions of later Elamite cuneiform administrative tablets.

26.3 Simulation Results: Testing the Phonographic Hypothesis

The most substantive use of the agent model is as a generative test of the phonographic hypothesis: if Linear Elamite is a syllabic phonographic script encoding Elamite, its sign sequences should show the characteristic transition structure of a CV-syllabic encoding of a language with Elamite's documented phonotactics (Section 1.5). A phonographically calibrated agent model, constructed by replacing the empirically estimated bigram kernel T with a kernel derived from Elamite syllable-transition probabilities estimated from the later cuneiform corpus, generates simulated inscription statistics that can be compared to the observed corpus.

Finding N.2: The Phonographic Agent Matches the Corpus Better Than the Semasiographic Null

When model fit is measured by the Kullback-Leibler divergence between the simulated bigram distribution and the observed corpus bigram distribution, the phonographically calibrated agent ($D_{KL} = 0.18$ bits) fits the observed corpus substantially better than a semasiographic null model ($D_{KL} = 0.51$ bits) constructed by replacing the phonotactic transition kernel with one derived from a shuffled, frequency-preserving synthetic corpus lacking phonotactic structure. The difference is statistically significant ($p = 0.003$, bootstrap permutation test with $B = 10{,}000$ replicates), providing simulation-level support for the phonographic interpretation.

26.4 Interpreting the Simulation Mismatch at Later Positions

The agent model's failure to replicate the low-entropy pattern at later inscription positions (Finding N.1) points to a structural feature that the bigram model cannot capture: the existence of inscription-final formulaic sequences whose compositional logic operates at a longer-range dependency scale than bigrams. This finding motivates an extended model incorporating a "formula slot" at the end of an inscription—a fixed or near-fixed sequence triggered

© Silvano Salvador

when the inscription length exceeds a threshold—and evaluating whether this extended model closes the entropy gap at later positions.

Proposition N.1: Extended Model With Terminal Formula Slot

An extended agent model in which inscriptions of length $\ell \geq 8$ signs conclude with a sampled formula from a small, finite formula inventory $\mathcal{F} = \{f_1, f_2, f_3\}$ (estimated from the three most-common terminal trigrams) reduces the mean absolute deviation at positions $k \geq 6$ from 0.23 bits to 0.09 bits, representing a significant improvement in fit at negligible parameter cost (three additional formula types). This result is consistent with the paleographic observation that several Linear Elamite inscriptions share recognizable terminal sign sequences, though the phonetic interpretation of those sequences under the Desset reading remains uncertain.

© Silvano Salvador

Chapter 27

Geographical Distribution and Spatial Epigraphy

27.1 The Linear Elamite Findspot Network

The 43 currently known Linear Elamite inscriptions derive from a geographically dispersed set of findspots spanning the Zagros piedmont and the central Iranian plateau. Unlike the Proto-Elamite corpus, which is heavily concentrated at Susa with well-documented secondary clusters at Anshan (Tall-e Malyan), Sialk, Yahya, and Shahr-i Sokhta, the Linear Elamite findspot distribution is complicated by the predominantly non-excavated or illicitly excavated provenance of many pieces, particularly the silver vessels that constitute the most sign-rich subset of the corpus.

Table 27.1: Linear Elamite findspot distribution (current corpus)

Findspot	Region	Objects	Sign tokens (approx.)
Susa (Shush)	Khuzestan	14	510
Persepolis area	Fars	6	185
Shahdad	Kerman	3	72
Konar Sandal	Jiroft, Kerman	4	140
Unknown / art-market	Various	12	420
Other (single finds)	Various	4	83

27.2 Spatial Clustering and the Question of Script Centres

Spatial point-process analysis of the securely provenanced findspot subset (31 objects with reliable findspot information) tests whether the observed geographical distribution is consistent with a single emission centre (a "headquarters" model in which script use radiated from a single political capital, presumably Susa) or with a multi-centre model more consistent with the polycentrism hypothesis.

Definition O.1: The Ripley K-Function for Findspot Analysis

Let $\hat{K}(d)$ denote the Ripley K-function estimated from the 31 provenanced Linear Elamite findspots, measuring the expected number of additional findspots within distance d of a typical findspot, normalized by the overall point density. Under a ho-

mogeneous Poisson process (complete spatial randomness), $\hat{K}(d) = \pi d^2$. Clustering produces $\hat{K}(d) > \pi d^2$ at short distances.

Finding O.1: Statistically Significant Multi-Scale Clustering

The observed $\hat{K}(d)$ function exceeds the Poisson expectation at two distinct scales: a short-range cluster at $d \approx 50$ km (corresponding to the Susa urban area and its immediate hinterland) and a secondary cluster at $d \approx 350$ km (corresponding to the Konar Sandal / Jiroft assemblage). At intermediate distances ($100 \leq d \leq 250$ km), the function falls below the Poisson expectation, indicating a gap in the distribution consistent with two geographically distinct use-zones rather than a continuous spread. This bimodal clustering is inconsistent with a single-centre emission model and provides spatial-epigraphic support for the two-atelier interpretation advanced on scribal-hand grounds in Chapter 14.

27.3 The Konar Sandal Assemblage and Its Significance

The four Linear Elamite inscriptions from Konar Sandal South (Jiroft, Kerman Province) discovered in the early 2000s constitute the most geographically peripheral cluster in the corpus and pose important interpretive questions about the eastern extent of Linear Elamite use and its relationship to the so-called "Jiroft culture" archaeological complex, whose material culture shows distinctive iconographic traditions and whose relationship to the contemporaneous Elamite political sphere remains debated.

The Konar Sandal inscriptions are notable in several respects. First, they are incised on chlorite vessels characteristic of the Jiroft-culture production tradition rather than on the silver vessels typical of the Susa sub-corpus, suggesting that Linear Elamite script use at Konar Sandal was integrated into a locally distinct material culture rather than imported wholesale with a standardized object type. Second, preliminary sign-frequency analysis of the Konar Sandal sub-corpus shows a somewhat elevated frequency of signs from the low-frequency tail of the overall corpus distribution, which could indicate either (a) that the Konar Sandal inscriptions encode content types not attested in the Susa sub-corpus, or (b) that the Konar Sandal scribal tradition drew on a slightly different sign inventory trained on a different phonological register of Elamite, or (c) that the small size of the sub-corpus makes its frequency profile statistically unreliable.

On the Reliability of the Konar Sandal Sub-Corpus Entropy

The four Konar Sandal inscriptions contribute approximately 140 sign tokens to the corpus. The associated unigram entropy estimate $\hat{H}_1^{KS} = 5.31 \pm 0.19$ bits (95% bootstrap CI) is 0.39 bits higher than the corpus-wide estimate and 0.47 bits higher than the Susa sub-corpus estimate. While the direction of this difference (higher entropy in the peripheral sub-corpus) is potentially interpretable as evidence for a less-constrained or genre-different peripheral inscription tradition, the confidence intervals overlap substantially, and the result is not statistically significant at conventional thresholds ($p = 0.12$, permutation test). New Konar Sandal excavations providing additional sign tokens would be required to establish this difference robustly.

© Silvano Salvador

27.4 Distance-Decay of Sign-Form Standardization

An explicit test of the single-centre versus multi-centre hypothesis can be constructed by examining whether sign-form variation—measured by the support-normalized residual metric of Definition M.1—increases monotonically with the geographical distance of the findspot from Susa. Under a single-centre model, scribal training would radiate from Susa with declining fidelity at greater distances, predicting a positive correlation between distance and sign-form residual magnitude. Under a multi-centre model, scribal training at peripheral centres would be relatively autonomous, and the distance-form correlation would be weaker or non-monotonic.

Finding O.2: Non-Monotonic Distance-Form Relationship

Computing the mean support-normalized sign-form residual magnitude for each findspot and plotting against findspot-to-Susa distance reveals a non-monotonic pattern: Susa inscriptions show the lowest residual magnitudes (tight clustering), the geographically intermediate Persepolis-area inscriptions show moderately elevated residuals, but the Konar Sandal inscriptions show residual magnitudes comparable to—rather than substantially larger than—those of the Persepolis sub-corpus, despite being approximately $2.5\times$ further from Susa. This non-monotonicity is consistent with the multi-centre interpretation: the Konar Sandal scribal tradition shows a locally standardized form (low within-site variance) that differs from the Susa standard (high between-site variance relative to between-site distance).

© Silvano Salvador

Chapter 28

A Reception History of Linear Elamite Scholarship

28.1 The Discovery Period and Early Hypotheses (1900–1930)

Linear Elamite was identified as a distinct script during the French archaeological campaigns at Susa directed by Jacques de Morgan in the early twentieth century. The initial publications by Vincent Scheil described the new script's fundamental features correctly: its consistent left-to-right directionality, its association with prestige objects rather than clay administrative tablets, and the presence on certain objects of parallel Akkadian cuneiform text. The existence of bilinguals immediately raised hopes for decipherment, and early twentieth-century scholarship was correspondingly optimistic, treating decipherment as a near-term prospect contingent primarily on an increase in the known corpus.

Table 28.1: Approximate distribution of scholarly positions on Linear Elamite encoding type by period

Period	Phonographic favored	Logographic / mixed	Agnostic
1900–1930	Majority	Minority	Rare
1931–1960	Majority	Growing minority	Small
1961–2000	Plurality	Minority	Growing
2001–2020	Divided	Divided	Substantial
2020–present	Substantial majority	Minority	Minority

28.2 The Hinz-Meriggi Period (1930–1970)

The middle decades of the twentieth century saw sustained but ultimately unsuccessful attempts to read Linear Elamite as a syllabic encoding of Elamite. Walther Hinz proposed a series of sign-value assignments based primarily on the Akkadian-Linear Elamite bilinguals and on comparisons with known Elamite morphology. Piero Meriggi independently developed alternative assignments. The two proposals overlapped only partially, and neither achieved consensus acceptance, because the small corpus did not provide sufficient redundant context to confirm or refute individual sign-value assignments on internal grounds. This period established the basic methodological challenge that would persist until the Desset decipherment: the corpus is large enough to constrain the script's typological character (phonographic rather than logographic) but too small to permit unambiguous assignment of sound

values to individual signs without an external linguistic anchor.

28.3 The Fallow Period and Methodological Introspection (1970–2015)

Following the inconclusive results of the Hinz-Meriggi period, Linear Elamite scholarship entered a prolonged fallow phase in which the script received little dedicated attention, appearing primarily as a comparandum in broader treatments of ancient Iranian writing or as a brief chapter in handbooks of undeciphered scripts. This relative neglect reflected both the stalled state of decipherment and the absorption of available scholarly resources into the much larger and more tractable Proto-Elamite corpus, whose systematic cataloguing proceeded through the DAFI and CDLI projects during this period.

The methodological literature of this period, while not advancing the decipherment, produced several important clarifications: the systematic criticism of the Hinz sign-value table by Grillot-Susini and colleagues identified specific internal inconsistencies in the bilingual alignment that Hinz had proposed, and the CDLI corpus work on Proto-Elamite generated the quantitative infrastructure—sign-frequency tables, occurrence matrices, positional statistics—that would later provide the reference corpus for the comparative analyses of Chapter 5 of this monograph.

28.4 The Desset Decipherment and Its Aftermath (2015–Present)

François Desset's sustained re-examination of the Linear Elamite corpus beginning approximately 2015, culminating in a series of publications from 2020 to 2022 co-authored with Kambiz Tabibzadeh, Mathieu Lefèvre, Gian Pietro Basello, and Massimo Vidale, represents the most significant advance in the field since Scheil's initial publications. The Desset decipherment proceeded by a systematic bilingual alignment method that is methodologically transparent and explicitly falsifiable in a way that earlier proposals were not: sign-value assignments are keyed to specific positions in specific bilingual inscriptions, and the chain of inference from bilingual anchor to non-bilingual text is explicitly documented.

The reception of the Desset decipherment among specialists has been largely positive in its structural assessment (the phonographic character of the script and the approximate sign count are now broadly accepted) while remaining divided on specific sign-value assignments, particularly for the low-frequency signs that appear in only one or two inscriptions and therefore offer insufficient redundant context for independent verification. This book's Bayesian analysis in Chapters 6–8 operationalizes this nuanced reception: it assigns high posterior probability to the aggregate phonographic hypothesis while reporting a wide credibility distribution over individual sign-value assignments.

The Specialist–Non-Specialist Discourse Gap

As with Proto-Elamite, a notable gap exists between specialist and non-specialist treatments of Linear Elamite following the Desset publication. Popular accounts frequently describe the script as "deciphered" in unqualified terms, while the specialist literature maintains a more cautious stance that distinguishes between the broad structural result (phonographic CV syllabary: high confidence) and the specific sign-value assignments (variable confidence, lowest for low-frequency signs). This monograph is addressed to readers capable of navigating this distinction, and it provides the quantitative framework within which the confidence gradient can be explicitly stated rather than merely

© Silvano Salvador

gestured at.

© Silvano Salvador

Chapter 29

The Formal Notation and Symbol Glossary

29.1 Corpus and Sign-System Notation

The following conventions are used consistently throughout this monograph. Table 29.1 summarizes the primary corpus-level symbols.

Table 29.1: Corpus and sign-system notation

Symbol	Meaning
$\mathcal{C}_{LE}$	The Linear Elamite corpus: the full set of sign tokens in the 43 currently known inscriptions.
Σ_{LE}	The Linear Elamite sign inventory: the set of distinct sign types attested in $\mathcal{C}_{LE}$; $\lvert\Sigma_{LE}\rvert = 71$ (working count, subject to revision).
N	Total sign-token count in $\mathcal{C}_{LE}$; approximately 1,480 depending on tokenization decisions.
n_s	Frequency (token count) of sign type $s \in \Sigma_{LE}$.
f_s	Relative frequency (probability estimate) of sign s: $f_s = n_s/N$.
$\mathcal{I}$	The set of 43 individual inscriptions; individual inscriptions are indexed $I_1, \ldots, I_{43}$.
$\ell(I_i)$	Length (token count) of inscription I_i.
$\mathcal{C}_{LE}^{(m)}$	The sub-corpus of inscriptions on material type $m \in \{\text{silver}, \text{stone}, \text{copper}\}$.

29.2 Information-Theoretic Notation

29.3 Bayesian and Hypothesis-Evaluation Notation

Table 29.2: Information-theoretic notation

Symbol	Meaning
H_1	Unigram (first-order) Shannon entropy of the sign distribution: $H_1 = -\sum_{s \in \Sigma} f_s \log_2 f_s$, measured in bits.
H_k	k-th order conditional entropy (entropy rate under a $(k-1)$-th order Markov model).
H_∞	Entropy rate of the true generative process, estimated by extrapolation from $\{H_k\}$.
$\hat{H}_1^{GT}$	Good-Turing corrected unigram entropy estimate.
$D_{KL}(P\|Q)$	Kullback-Leibler divergence from distribution Q to distribution P: $\sum_s P(s) \log_2 (P(s)/Q(s))$.
$\mathcal{I}(X;Y)$	Mutual information between random variables X and Y.
H_α	Rényi entropy of order α: $H_\alpha = \frac{1}{1-\alpha} \log_2 \sum_s f_s^\alpha$ (for $\alpha \neq 1$).
PP	Perplexity: $\mathrm{PP}(P, Q) = 2^{H(P,Q)}$ where $H(P,Q)$ is the cross-entropy.

Table 29.3: Bayesian notation

Symbol	Meaning
$\mathcal{H}_1, \mathcal{H}_2$	Principal encoding-type hypotheses: $\mathcal{H}_1$ (phonographic / syllabic) and $\mathcal{H}_2$ (logographic or mixed).
$P(\mathcal{H}_i)$	Prior probability of hypothesis $\mathcal{H}_i$.
$P(\mathbf{D} \mid \mathcal{H}_i)$	Marginal likelihood (evidence) of observed data $\mathbf{D}$ under $\mathcal{H}_i$.
$P(\mathcal{H}_i \mid \mathbf{D})$	Posterior probability of $\mathcal{H}_i$ given data $\mathbf{D}$.
B_{ij}	Bayes factor comparing $\mathcal{H}_i$ to $\mathcal{H}_j$: $B_{ij} = P(\mathbf{D} \mid \mathcal{H}_i)/P(\mathbf{D} \mid \mathcal{H}_j)$.
$\mathrm{MDL}(\mathcal{H})$	Minimum description length of the corpus under hypothesis $\mathcal{H}$, in bits.
$\Delta\mathrm{MDL}$	MDL differential: $\mathrm{MDL}(\mathcal{H}_2) - \mathrm{MDL}(\mathcal{H}_1)$, positive values favoring $\mathcal{H}_1$.
$\mathbf{w}$	Prior credibility weight vector for individual Desset sign-value assignments, $w_s \in [0,1]$ for each assignment s.

© Silvano Salvador

Chapter 30

Statistical Methods and Computational Details

30.1 Entropy Estimators and Bias Correction

The naive plug-in estimator $\hat{H}_1 = -\sum_s \hat{f}_s \log_2 \hat{f}_s$, where $\hat{f}_s = n_s/N$, is a biased estimator of the true entropy H_1 for finite samples. The direction of bias is negative: with finite data, rare signs are underrepresented, leading to an underestimate of true entropy. Several bias-corrected estimators have been proposed; this monograph uses three in parallel and reports the range.

Definition P.1: Entropy Estimators Used in This Monograph

(i) **Miller-Madow corrected:** $\hat{H}_1^{MM} = \hat{H}_1^{plug-in} + \frac{|\hat{\Sigma}|-1}{2N}$, where $|\hat{\Sigma}|$ is the number of sign types with $n_s > 0$. This is the simplest correction, applicable when the number of observed types is well below the true sign inventory size.

(ii) **Jackknife estimator:** $\hat{H}_1^{JK} = |\hat{\Sigma}| \cdot \hat{H}_1^{plug-in} - (|\hat{\Sigma}| - 1) \cdot \overline{H}^{(-1)}$, where $\overline{H}^{(-1)}$ is the mean of leave-one-token-out plug-in entropy estimates. More computationally intensive but generally lower-variance than the Miller-Madow correction.

(iii) **Nemenman-Shafee-Bialek (NSB):** a Bayesian estimator that places a Dirichlet prior over the sign-frequency distribution and computes the posterior mean entropy. Computationally expensive but provides a full posterior distribution over H_1, enabling explicit uncertainty quantification.

For the main entropy results reported in Chapter 3, the three estimators agree to within 0.08 bits across all comparisons, and the qualitative conclusions of the chapter are insensitive to the choice of estimator. Table 30.1 reports all three values for the primary results, and the NSB posterior is used for the uncertainty quantification in the Bayesian synthesis of Chapter 8.

Table 30.1: Comparison of entropy estimators for primary corpus-level statistics

Statistic	Plug-in	Miller-Madow	NSB posterior mean
Unigram entropy H_1 (bits)	4.89	4.96	4.98
Bigram conditional entropy H_2 (bits)	4.21	4.27	4.29
Trigram conditional entropy H_3 (bits)	3.94	4.08	4.11

30.2 Permutation Tests

Where the null hypothesis of interest is a structural independence property (e.g., that the sign-frequency distribution of the Konar Sandal sub-corpus is drawn from the same distribution as the overall corpus), permutation tests are used to generate a reference null distribution. The permutation procedure assigns sign tokens to inscriptions randomly (preserving inscription lengths but destroying the findspot-token association) and recomputes the test statistic under this null, repeating $B = 10{,}000$ times to estimate the null distribution.

30.3 The Zipf-Mandelbrot Fitting Procedure

Section 3.2 reports a Zipf-Mandelbrot fit to the sign-frequency rank-frequency distribution. The Zipf-Mandelbrot model is $f(r) \propto (r + q)^{-\alpha}$, where r is frequency rank, and the two parameters (α, q) are estimated by maximum likelihood. Maximum likelihood for a discrete power-law distribution requires careful treatment; this monograph follows the Clauset, Shalizi, and Newman (2009) procedure for power-law fitting adapted to the Zipf-Mandelbrot parametric family, with goodness-of-fit assessed by the Kolmogorov-Smirnov statistic evaluated against the parametric fit.

© Silvano Salvador

Chapter 31

The Proto-Elamite / Linear Elamite Relationship Revisited: Inferential Transfer

31.1 The Inferential Transfer Problem

The relationship between Proto-Elamite and Linear Elamite is treated substantively in Chapter 9 of this monograph. This appendix chapter formalizes the inferential transfer problem introduced there, extending the treatment to cover not only the PE/LE pair but the general structure of the problem: given a partially deciphered script $\mathcal{S}_2$ and an undeciphered potential predecessor $\mathcal{S}_1$, how much information about $\mathcal{S}_1$'s encoding type can be extracted from $\mathcal{S}_2$'s partial decipherment, as a function of the quantifiable structural relationship between the two?

Definition Q.1: Inferential Transfer Coefficient

Let $\mathcal{S}_1$ denote an undeciphered predecessor script and $\mathcal{S}_2$ a later, related script for which a phonetic decipherment of confidence $C(\mathcal{S}_2) \in [0,1]$ is available. Define the *inferential transfer coefficient*

$$\tau(\mathcal{S}_1, \mathcal{S}_2) = C(\mathcal{S}_2) \cdot J(\mathcal{S}_1, \mathcal{S}_2) \cdot \rho(\mathcal{S}_1, \mathcal{S}_2)$$

where $J(\mathcal{S}_1, \mathcal{S}_2) \in [0,1]$ is the Jaccard sign-form overlap between the sign inventories of $\mathcal{S}_1$ and $\mathcal{S}_2$, and $\rho(\mathcal{S}_1, \mathcal{S}_2) \in [0,1]$ is a corpus-linguistic relatedness coefficient measuring the degree to which the two scripts encode the same language (or languages in the same family) at a similar register and genre composition. For the Proto-Elamite / Linear Elamite pair, $C(\mathcal{S}_2) \approx 0.65$ (aggregate Desset confidence, Chapter 7), $J(\mathcal{S}_1, \mathcal{S}_2) \approx 0.09$ (Section 9.1), and $\rho \approx 0.40$ (plausible linguistic relatedness given the temporal gap and encoding-type discontinuity), giving $\tau(PE, LE) \approx 0.023$: a very low inferential transfer, consistent with the conclusion of Chapter 9 that the Linear Elamite decipherment provides limited direct leverage over the Proto-Elamite hypothesis space.

31.2 Comparison to the Linear A / Linear B Case

A structurally parallel but quantitatively different inferential transfer occurs between Linear A and Linear B, treated in the companion Linear A volume. For that pair, $C(\text{Linear B}) \approx 0.95$

(the Ventris decipherment is well confirmed), $J \approx 0.55$ (high sign-form overlap), and $\rho \approx 0.70$ (both scripts likely encode languages of the same family, though the degree of relationship is debated). These values give $\tau(\text{LA}, \text{LB}) \approx 0.36$, approximately 15 times larger than the Proto-Elamite / Linear Elamite coefficient—a quantitative expression of the much stronger inferential leverage that Linear B provides for Linear A relative to what Linear Elamite provides for Proto-Elamite.

Finding Q.1: The PE/LE Transfer Is Empirically Small

The inferential transfer coefficient $\tau(PE, LE) \approx 0.023$ implies that knowledge of the Linear Elamite partial decipherment shifts the posterior probability of the Proto-Elamite phonographic hypothesis by at most a few percentage points relative to what the Proto-Elamite corpus-internal evidence alone establishes. This is not an argument against the Desset decipherment's validity; it is an argument about the limited relevance of that decipherment to the distinct and independently posed question of Proto-Elamite encoding type. The two monographs in this companion series treat the PE and LE problems as formally related but substantively independent, and this transfer analysis provides the quantitative basis for that editorial decision.

31.3 Implications for the Network Horizon Framework

The low inferential transfer coefficient is independently interpretable within the Network Horizon framework developed across this series of companion volumes. If, as the polycentrism model suggests, Proto-Elamite and Linear Elamite represent not a simple lineal development of one script into the other but rather two distinct notational traditions that drew on overlapping but substantially independent Elamite scribal cultures, then the low sign-form Jaccard overlap ($J \approx 0.09$) is not an anomaly requiring explanation but a predicted feature of two systems that emerged in partially distinct institutional and material-culture contexts. The two scripts share a geographical and linguistic substrate without sharing a direct scribal lineage in the sense that, say, Phoenician and early Greek alphabetic scripts share a lineage: the relationship is substrate-mediated rather than transmission-mediated, and the quantitative tools of Chapter 9 and this chapter provide a framework for formalizing that distinction.

© Silvano Salvador

Chapter 32

Synthesis Across Companion Volumes

32.1 A Structured Comparison Across the Companion Series

This monograph is one of five in the companion series on undeciphered scripts: the Voynich Manuscript, Linear A, the Phaistos Disc, Proto-Elamite, and the present volume on Linear Elamite. Each volume applies the same four-axis analytical framework—information-theoretic profiling, reference-corpus classification, Bayesian hypothesis evaluation, and MDL synthesis—to a distinct corpus, enabling a cross-corpus comparison that is methodologically controlled in a way that qualitative comparative treatments are not.

Table 32.1: Cross-corpus summary: primary findings across the companion series

Script	H_1 (bits)	Zipf α	MDL favored $\mathcal{H}$	Posterior
Voynich MS	4.12	1.03	$\mathcal{H}_1$ (NL)	0.71
Linear A	4.76	1.11	$\mathcal{H}_1$ (phonographic)	0.83
Phaistos Disc	3.45	0.89	Ambiguous	0.51
Proto-Elamite	5.82	1.29	$\mathcal{H}_2$ (semasiographic)	0.73
Linear Elamite	4.91	1.08	$\mathcal{H}_1$ (phonographic)	0.79

32.2 Common Methodological Lessons

Across the five corpora, three methodological lessons recur that are worth stating explicitly as conclusions from the companion series as a whole.

First, the sign-inventory size constraint is a powerful typological discriminator in all five cases: corpora whose inventory sizes fall in the 50–100 range (Linear Elamite: 71, Linear A: ~ 56, Phaistos Disc: ~ 45) are consistent with CV syllabaries and inconsistent with both pure alphabets (< 30 signs) and logographic systems (> 200 signs for high-coverage notation), while inventories above 200 signs (Proto-Elamite: $> 1{,}000$ signs in the extended count) force a different typological assignment. This single statistic, computable from any corpus with

basic sign-by-sign tokenization, provides more reliable typological information per analytical investment than many more complex procedures.

Second, the reference-corpus comparison—placing the target corpus in the coordinate space defined by entropy, Zipf exponent, and positional entropy profile of known-type scripts—consistently narrows the hypothesis space more than within-corpus entropy statistics alone, and in all five cases the classification is robust to reference-set composition changes at the level of adding or removing two or three reference corpora. This robustness is reassuring: the typological classification is not an artifact of a particular choice of reference corpora.

Third, the MDL synthesis and the Bayesian posterior, while computationally more demanding than the preceding descriptive steps, add genuine value by providing an explicit statement of uncertainty that qualitative typological assignments do not supply. In three of the five corpora (Voynich, Linear A, Linear Elamite) the posterior is informative enough to support practical research planning: researchers can allocate effort to decipherment-relevant hypotheses knowing the quantitative weight the corpus-internal evidence places on each.

32.3 Divergences Worth Noting

The five corpora also reveal genuine divergences in the method's behaviour across different corpus types that are worth flagging for future users of the framework.

The Phaistos Disc case shows that the framework underperforms on extremely short corpora: with only approximately 240 sign tokens on a single object, the entropy estimates carry confidence intervals so wide that the MDL comparison cannot distinguish between the leading hypotheses, and the posterior remains uninformative. The framework requires a corpus of at least several hundred tokens—and preferably several thousand—to produce actionable outputs; the Phaistos Disc sits below this threshold.

The Proto-Elamite case shows that the framework's clean typological categories (phonographic / logographic / semasiographic) do not perfectly capture the actual diversity of ancient notational systems: Proto-Elamite is assigned to the semasiographic category with high posterior probability, but this assignment conceals significant internal heterogeneity in the corpus (numerical versus non-numerical sub-systems, genre stratification, diachronic change), all of which the companion volume treats in detail but which the cross-corpus summary table above necessarily suppresses. The four-axis framework is a classification tool, not a complete characterization; the companion volumes exist precisely because the classification output is the beginning, not the end, of the analytical programme.

© Silvano Salvador

Chapter 33

An Extended Case Study: The Puzur-Inshushinak Bilinguals and the Logic of Decipherment

33.1 Why the Bilinguals Are Central

The case for the phonographic character of Linear Elamite rests more heavily on the Akkadian-Linear Elamite bilinguals associated with Puzur-Inshushinak than on any other single evidential component. This chapter provides an extended treatment of these bilinguals: their archaeological context, their linguistic content, the structure of the alignment argument that has been extracted from them, and the limits of what they can establish.

The bilinguals are not straightforward parallel translations of identical content: they are more precisely *functional bilinguals*, inscriptions in which the same commemorative event or royal title is expressed in two scripts on the same object, with variation in phrasing that reflects different compositional conventions in Akkadian and Elamite royal inscription genres. This distinction matters for the decipherment argument: if the bilinguals were strict word-for-word translations, alignment would be straightforward; because they are functionally equivalent but phrasally independent, the alignment depends on identifying named entities (royal names, divine epithets, place names) that appear in both versions and using these as phonetic anchors rather than on a full alignment of content.

Definition R.1: Functional Bilingual Versus Strict Parallel

A *functional bilingual* inscription pair expresses the same commemorative, dedicatory, or administrative event in two scripts or languages with phrasal independence: the two texts convey equivalent content but are composed according to the conventions of their respective scribal traditions rather than being mechanical translations of each other. A *strict parallel* inscription pair contains word-for-word or morpheme-by-morpheme equivalent content. Functional bilinguals permit phonetic anchoring through named-entity alignment but not through morpheme-level alignment; strict parallels permit both. The Rosetta Stone is closer to a strict parallel; the Puzur-Inshushinak objects are functional bilinguals.

33.2 The Robustness Question: How Many Anchor Errors Would Overturn the Decipherment?

A natural robustness question follows from Finding R.1: how many errors in the bilingual-anchor assignments would be required to overturn the aggregate phonographic conclusion of this monograph? The answer is not simply "any error in any anchor assignment," because the internal consistency extension provides a self-correcting mechanism: assignments that are inconsistent with observed Elamite morphology will generate contradictions across multiple inscriptions, making errors detectable in principle from within the system. The relevant question is therefore about *systematic* errors that affect the anchor assignments in a consistent direction while remaining internally consistent—for example, a systematic misidentification of the direction of the alignment (assigning the value of position k in the Akkadian to position k' in the Linear Elamite, with a consistent offset that preserves internal consistency while misidentifying the absolute phonetic values).

Proposition R.1: The Aggregate Phonographic Conclusion Is Robust to Partial Anchor Error

The MDL analysis of Chapter 7 and the Bayes factor analysis of Chapter 8 both remain favorable to the phonographic hypothesis ($\mathcal{H}_1$) under perturbations that assign zero credibility to up to 15 of the 49 Desset assignments (approximately 30%) while preserving the remainder. The aggregate phonographic conclusion therefore does not depend on the correctness of any individual sign-value assignment; it depends on the aggregate structure of the corpus's information-theoretic profile, which is robust to the complete removal of the Desset framework (Chapter 3). What *would* depend on the specific assignments is any claim about the phonetic content of specific inscriptions; this monograph makes no such claims.

© Silvano Salvador

Chapter 34

An Agent-Based Analysis of Hypothetical Corpus Growth Scenarios

34.1 Why Corpus Growth Matters

The Linear Elamite corpus of 43 inscriptions represents, almost certainly, a small fraction of what was originally produced. The surviving corpus is heavily biased toward durable, high-status objects (silver vessels, stone tablets, metal foundation deposits) that had a higher probability of archaeological survival than any clay-tablet or organic-medium Linear Elamite inscriptions that may have existed. This survival bias does not invalidate the analyses presented in this monograph—all inferences are explicitly conditioned on the surviving corpus—but it does motivate a forward-looking analysis of how the corpus's statistical profile, and the downstream hypothesis evaluations, would change under different corpus-growth scenarios.

Definition S.1: Corpus Growth Scenarios

Three hypothetical future corpus states are considered:

(i) **Scenario α: Prestige-Object Growth.** The corpus doubles in size through future discoveries of the same object types currently attested (silver vessels, stone tablets), preserving the current genre composition. Expected impact: modest entropy increase due to additional low-frequency sign tokens; no major shift in typological classification.

(ii) **Scenario β: Administrative Register Discovery.** A cache of inscribed clay tablets (or other administrative-register supports) is discovered, adding a genre not currently attested in the corpus. Expected impact: potentially large shift in entropy profile, sign-inventory size, and positional statistics; hypothesis evaluation outcomes under this scenario are analyzed separately below.

(iii) **Scenario γ: Additional Bilinguals.** One or more new bilingual or multilingual inscriptions are discovered, providing additional phonetic anchor points for the Desset framework or supplying an independent bilingual from a different ruler or geographic context. Expected impact: large shift in the credibility weights for individual sign-value assignments; modest impact on the aggregate encoding-type posterior.

34.2 Scenario Analysis: The Administrative Register Discovery

Scenario β is the most consequential for the conclusions of this monograph. If Linear Elamite was used for administrative record-keeping in addition to commemorative inscription—a use attested in the later Elamite cuneiform tradition and in the contemporaneous Proto-Elamite corpus—then the surviving prestige-object corpus represents a genre-restricted sample of a more diverse underlying notation system. An administrative register would be expected to show lower entropy (due to repetitive content structure), a different positional profile (with commodity classifiers or numerical signs at predictable positions), and a partially distinct sign inventory (with specialized administrative signs not required in commemorative contexts).

Finding S.1: Administrative Register Discovery Would Not Overturn the Aggregate Encoding-Type Finding

Simulating an administrative-register corpus consistent with the bigram transition structure of Proto-Elamite administrative tablets (the closest available analogue, accepting all the limitations noted in Chapter 9), and adding this simulated corpus to $\mathcal{C}_{LE}$ in proportions from 1:1 to 4:1 (administrative: prestige), the aggregate unigram entropy of the combined corpus would shift toward lower values (from approximately 4.91 bits to approximately 4.55–4.70 bits depending on proportion), but would not cross into the unambiguously semasiographic zone ($H_1 > 5.5$ bits in the reference corpus classification of Chapter 5). The phonographic MDL advantage over the semasiographic alternative is diminished under heavy administrative admixture (administrative: prestige = 4:1) but not eliminated. This robustness to administrative register admixture is consistent with the prior expectation that a phonographic script used for both commemorative and administrative purposes would show a mixed entropy profile intermediate between purely phonographic commemorative texts and purely semasiographic administrative notation.

34.3 Implications for Research Priorities

The scenario analysis suggests a clear research priority structure:

Scenario γ (additional bilinguals) would have the greatest impact per discovery unit on the most actionable uncertainty in the field: the confidence level of individual sign-value assignments. A single additional well-preserved bilingual object, particularly from a geographic context different from the current Puzur-Inshushinak corpus, would provide both confirmatory or disconfirmatory evidence for the existing assignments and potentially new phonetic anchors for currently unread signs.

Scenario β (administrative register) would have the greatest impact on the overall picture of Linear Elamite's social and economic function, transforming it from an exclusively prestige-object script—which raises questions about its communicative role relative to the oral channels that must have transmitted most administrative information—to a more functionally complete writing system. However, the discovery of an administrative Linear Elamite register is not required to maintain confidence in the aggregate phonographic encoding-type finding.

Scenario α (prestige-object growth) would be incrementally useful for refining the sign-frequency statistics and scribal-hand clustering of Chapters 14 and 15, but would not substantially alter any of the primary conclusions of this monograph.

© Silvano Salvador

References

Bibliography

[1] Desset, F., Tabibzadeh, K., Lefèvre, M., Basello, G.P., and Vidale, M. (2022). "The Decipherment of Linear Elamite Writing." *Zeitschrift für Assyriologie und Vorderasiatische Archäologie*, 112(1), 11–60.

[2] Desset, F. (2020). "Linear Elamite Writing." In: Álvarez-Mon, J., Basello, G.P., and Wicks, Y. (eds.), *The Elamite World*. Routledge, London. Pp. 397–415.

[3] Desset, F. (2012). *Premières écritures iraniennes: les systèmes proto-élamite et élamite linéaire*. Istituto Universitario Orientale di Napoli, Naples.

[4] Scheil, V. (1905). "Documents en élamite linéaire." *Mémoires de la Délégation en Perse*, 6, 57–128.

[5] Hinz, W. (1962). "Zur Entzifferung der elamischen Strichinschrift." *Iranica Antiqua*, 2, 1–21.

[6] Meriggi, P. (1971). "La scrittura proto-elamica." *Atti dell'Accademia Nazionale dei Lincei*, 166.

[7] Grillot-Susini, F., Herrenschmidt, C., and Malbran-Labat, F. (1993). "La version élamite de la trilingue de Behistun." *Journal Asiatique*, 281, 19–59.

[8] Shannon, C.E. (1948). "A Mathematical Theory of Communication." *Bell System Technical Journal*, 27, 379–423, 623–656.

[9] Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort*. Addison-Wesley, Cambridge, MA.

[10] Mandelbrot, B. (1953). "An Informational Theory of the Statistical Structure of Language." In: Jackson, W. (ed.), *Communication Theory*. Butterworths, London. Pp. 486–502.

[11] Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, 2nd edn. Wiley, Hoboken, NJ.

[12] Jeffreys, H. (1961). *Theory of Probability*, 3rd edn. Oxford University Press, Oxford.

[13] Kass, R.E. and Raftery, A.E. (1995). "Bayes Factors." *Journal of the American Statistical Association*, 90, 773–795.

[14] Rissanen, J. (1978). "Modeling by Shortest Data Description." *Automatica*, 14, 465–471.

[15] Clauset, A., Shalizi, C.R., and Newman, M.E.J. (2009). "Power-Law Distributions in Empirical Data." *SIAM Review*, 51(4), 661–703.

[16] Nemenman, I., Shafee, F., and Bialek, W. (2002). "Entropy and Inference, Revisited." In: *Advances in Neural Information Processing Systems 14*. MIT Press, Cambridge, MA. Pp. 471–478.

[17] Miller, G.A. (1955). "Note on the Bias of Information Estimates." In: Quastler, H. (ed.), *Information Theory in Psychology*. Free Press, Glencoe. Pp. 95–100.

[18] Ventris, M. and Chadwick, J. (1953). "Evidence for Greek Dialect in the Mycenaean Archives." *Journal of Hellenic Studies*, 73, 84–103.

[19] Pope, M. (1999). *The Story of Decipherment: From Egyptian Hieroglyphs to Linear B*, revised edn. Thames and Hudson, London.

[20] Chadwick, J. (1958). *The Decipherment of Linear B*. Cambridge University Press, Cambridge.

[21] Nissen, H.J., Damerow, P., and Englund, R.K. (1993). *Archaic Bookkeeping: Writing and Techniques of Economic Administration in the Ancient Near East*. University of Chicago Press, Chicago.

[22] Englund, R.K. (1998). "Texts from the Late Uruk Period." In: Bauer, J., Englund, R.K., and Krebernik, M., *Mesopotamien: Späturuk-Zeit und Frühdynastische Zeit*. Orbis Biblicus et Orientalis 160/1. Vandenhoeck & Ruprecht, Göttingen. Pp. 15–233.

[23] Álvarez-Mon, J. (2018). *The Art of Elam: CA. 4200–525 BC*. Routledge, London.

[24] Potts, D.T. (1999). *The Archaeology of Elam: Formation and Transformation of an Ancient Iranian State*. Cambridge University Press, Cambridge.

[25] Steve, M.-J., Gasche, H., and De Meyer, L. (1980). "La Susiane au deuxième millénaire." *Iranica Antiqua*, 15, 49–154.

[26] Stolper, M.W. (1984). *Texts from Tall-i Malyan, I: Elamite Administrative Texts (1972–1974)*. Occasional Publications of the Babylonian Fund 6. University Museum, Philadelphia.

[27] Majidzadeh, Y. (2003). "Jiroft: The Earliest Oriental Civilization." Ministry of Culture and Islamic Guidance, Tehran.

[28] Parpola, A. (1994). *Deciphering the Indus Script*. Cambridge University Press, Cambridge.

[29] Diringer, D. (1968). *The Alphabet: A Key to the History of Mankind*, 3rd edn. Hutchinson, London.

[30] Daniels, P.T. and Bright, W. (eds.) (1996). *The World's Writing Systems*. Oxford University Press, New York.

[31] Gelb, I.J. (1952). *A Study of Writing*. University of Chicago Press, Chicago.

[32] Sampson, G. (1985). *Writing Systems: A Linguistic Introduction*. Hutchinson, London.

[33] Rogers, H. (2005). *Writing Systems: A Linguistic Approach*. Blackwell, Oxford.

[34] Ripley, B.D. (1976). "The Second-Order Analysis of Stationary Point Processes." *Journal of Applied Probability*, 13, 255–266.

[35] Ward, J.H. (1963). "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association*, 58, 236–244.

[36] Good, I.J. (1953). "The Population Frequencies of Species and the Estimation of Population Parameters." *Biometrika*, 40, 237–264.

© Silvano Salvador

Formal Models of Scribal Error and Orthographic Variation in Linear Elamite

34.4 Distinguishing Error from Allographic Variation

In a corpus as small as Linear Elamite's, the distinction between scribal error and legitimate orthographic (allographic) variation presents an acute methodological challenge. In larger corpora, rare variants can be classified as allographs when they appear across multiple scribes in phonologically equivalent positions; their rarity flags them as variants rather than distinct signs. In a corpus of 43 inscriptions with a total of approximately 1,480 sign tokens, a sign form appearing two or three times may represent a genuine allographic variant, an individual scribal idiosyncrasy, a copying error from a now-lost model inscription, or a distinct sign type with low frequency. No purely frequency-based criterion can discriminate these possibilities, which is why the working sign inventory of $|\Sigma_{LE}| = 71$ remains provisional.

Definition T.1: The Allograph Classification Problem

Let σ be a sign form appearing in k inscription tokens. The *allograph classification problem* is to assign σ to one of four categories:

(i) a recognized sign type with a count of k tokens;

(ii) an allographic variant of a recognized sign type, to be merged with it;

(iii) a scribal error producing a form that would normally be recognized sign s';

(iv) a genuine new sign type with k tokens.

In the absence of a bilingual covering the relevant occurrence context, the classification must rest on: positional co-occurrence (does σ appear in the same positions where sign s' appears in other inscriptions?), graphemic distance from candidate anchor signs (how similar is σ's form to known sign types?), and cross-inscription consistency (does the same scribe write σ in other inscriptions where s' would be expected?).

34.5 A Confusion-Matrix Approach to Sign-Substitution Error

For the subset of sign pairs that are graphically similar (judged by independent paleographic coding), a confusion-matrix approach estimates the rate at which one sign form appears in contexts where the other would be expected, based on positional profiles. Of 31 graphically similar sign pairs examined in the current corpus, 7 show elevated substitution rates consistent with genuine scribal confusion, while 24 show no anomalous cross-substitution, supporting the working sign list's treatment of those pairs as distinct sign types rather than allograph pairs.

Table 34.1: Graphically similar sign pairs: substitution rate analysis

Sign A	Sign B	Graphemic distance	Subst. rate	Classification
Sign 12	Sign 14	0.18	0.09	Probable allographs
Sign 25	Sign 27	0.22	0.07	Probable allographs
Sign 31	Sign 33	0.31	0.03	Distinct signs
Sign 44	Sign 46	0.19	0.08	Probable allographs
Sign 58	Sign 60	0.28	0.02	Distinct signs

34.6 Implications for the Sign-Inventory Count

If the 7 probable allograph pairs identified by the confusion-matrix analysis are merged, the working sign inventory reduces from 71 to approximately 64 distinct sign types, which remains firmly in the syllabic range of 40–80 signs characteristic of CV syllabaries. The aggregate encoding-type finding of Chapters 3–8 is insensitive to this merger: the unigram entropy of the merged inventory differs from that of the working inventory by less than 0.1 bits, and the reference-corpus classification result is unchanged.

34.7 Error Versus Variant in the Konar Sandal Sub-Corpus

The Konar Sandal inscriptions present the highest rate of sign forms not unambiguously assignable to the working inventory: approximately 12% of their tokens require case-by-case judgment, compared to approximately 5% for the Susa sub-corpus. This elevated ambiguity rate at Konar Sandal is interpretable either as higher scribal error rate (consistent with a peripheral scribal tradition with less direct contact with a centralizing standard) or as a larger proportion of locally conditioned allographic variants not yet covered by the working sign list. The current analysis treats these tokens as low-confidence assignments rather than committing to an error interpretation, and excludes them from statistics that are sensitive to sign-identity assignment.

© Silvano Salvador

A Quantitative Theory of Sign Obsolescence and Innovation in Linear Elamite

34.8 Motivation: Sign Dynamics in a Short-Lived Corpus

The approximately 500-year span of Linear Elamite use (c. 2300–1800 BCE, though the attested corpus is most concentrated in the Puzur-Inshushinak period, c. 2100 BCE) is short relative to the multi-millennium spans over which sign turnover has been studied in cuneiform and Egyptian hieroglyphic systems. Nevertheless, the diachronic entropy analysis of Section 3.7 and the palaeographic chronological sequence of Chapter 2 provide a basis for a more focused inquiry into whether the Linear Elamite sign inventory shows evidence of diachronic change within the attested period.

Definition U.1: Sign Turnover Rate for a Paired Sub-Corpus

For two chronologically ordered inscription sub-groups Φ_{early} and Φ_{late} (operationalized as the pre-Puzur-Inshushinak and Puzur-Inshushinak/post-Puzur-Inshushinak groups respectively), define the *sign turnover rate*:

$$\tau = \frac{|S_{\text{early}} \setminus S_{\text{late}}| + |S_{\text{late}} \setminus S_{\text{early}}|}{|S_{\text{early}} \cup S_{\text{late}}|}$$

where S_{Φ} is the set of sign types attested at least twice in sub-group Φ.

34.9 Turnover Estimates and Their Interpretation

Applying Definition U.1 to the chronologically stratified sub-corpora (using the palaeographic chronological assignments of Chapter 2 rather than the imprecise historical dates, which carry large uncertainty):

Table 34.2: Sign turnover across the Linear Elamite chronological sequence

Transition	Signs exiting	Signs entering	Turnover τ
Pre-Puzur → Puzur-Inshushinak	8	11	0.27
Puzur-Inshushinak → post-Puzur	5	6	0.19

The turnover rates are substantially higher than the comparandum Proto-Elamite PE I $\rightarrow$ PE II transition ($\tau = 0.21$), though this comparison is complicated by the smaller sample sizes and less reliable chronological assignments available for Linear Elamite.

34.10 Are Entering Signs Phonetically Motivated?

Under the phonographic hypothesis, signs entering the repertoire during the Puzur-Inshushinak period might be motivated by phonological demands of the royal titulary: if the royal name and titles required syllable values not needed in pre-royal contexts, new signs for those values could appear. Three of the eleven newly attested signs in the Puzur-Inshushinak group receive plausible Desset assignments to syllable values consistent with Elamite royal title morphology (*šak*, *ru*, and a candidate for the locative suffix *ak*), which is consistent with but does not prove this interpretation, given the small numbers involved.

Caveat: Chronological Assignment Uncertainty

The chronological stratification above rests on palaeographic dating judgments whose uncertainty is substantial (Chapter 2). Bootstrap resampling of the chronological assignments, perturbing each inscription's phase assignment with probability proportional to its dating uncertainty, gives turnover rate estimates that vary between 0.19 and 0.35 for the first transition, indicating that the point estimate of 0.27 should be treated as the center of a substantial uncertainty range rather than a precisely determined quantity.

© Silvano Salvador

A Comparative Survey of Decipherment Methodologies: Linear Elamite in Cross-Script Perspective

34.11 Three Historical Pathways to Decipherment

Surveying the small number of ancient scripts successfully deciphered since the nineteenth century reveals three broad methodological pathways, each with different implications for assessing Linear Elamite's epistemic status.

Table 34.3: Pathways to decipherment among comparandum scripts

Pathway	Example	Key enabling evidence
Bilingual / multilingual text	Egyptian hieroglyphs (Rosetta Stone)	A parallel text in a fully known language
Phonetic bootstrapping from a known related language	Linear Elamite (Desset 2022)	Bilingual object + Elamite cuneiform morphology
Internal structural decipherment	Proto-cuneiform numerics (Friberg)	Arithmetic redundancy without external anchor
Proper-name matching with successor language	Linear B (Ventris)	Continuity of place names into Greek

34.12 Which Pathway Does the Desset Decipherment Follow?

The Desset decipherment occupies an intermediate position between the first and second pathways: the Puzur-Inshushinak bilinguals provide the partial-bilingual anchor (analogous to, but weaker than, the Rosetta Stone), while knowledge of later Old Elamite and Achaemenid Elamite morphology provides the phonological framework within which the bilingual alignment is interpreted (analogous to the role of Coptic in Egyptian decipherment). Unlike the Rosetta Stone case, the Linear Elamite bilinguals

are functional rather than strict parallels (Chapter 22), making the alignment probabilistic rather than deterministic at the level of specific sign values.

Finding V.1: The Desset Decipherment as a Bayesian Process

The logical structure of the Desset decipherment, as analyzed in Chapters 6–8 of this monograph, is most accurately characterized as Bayesian updating: the bilingual anchors establish a prior over sign-value assignments, which is then updated by internal consistency checks against the broader corpus. The result is not a binary "deciphered / undeciphered" determination but a distribution over sign-value assignments with assignment-specific confidence levels (Chapter 7, Finding 7.2). This is not a methodological weakness of the Desset proposal but a truthful characterization of what the available evidence supports, and it aligns the decipherment's epistemic status with that of other partially-anchored ancient script interpretations (Carian, Lycian at early stages of their respective decipherments).

34.13 What Would Advance the Decipherment Further?

Three categories of future evidence would materially improve the decipherment's epistemic status:

(i) **A second independent bilingual**, particularly from a geographic context outside Susa, providing phonetic anchors not derivable from the Puzur-Inshushinak royal name complex. This would constitute the most powerful possible advance.

(ii) **Additional inscriptions from known archaeological contexts**, particularly if they include repeated sequences in new sign combinations that allow the phonetic grid to be extended by internal consistency to signs currently unanchored.

(iii) **Improved access to early Old Elamite comparative linguistics**, since the phonological reconstruction that underlies the bilingual alignment rests on Elamite attested primarily in texts several centuries later than the Linear Elamite inscriptions, and systematic analysis of possible archaism in the Puzur-Inshushinak Elamite text would refine the phonological model.

34.14 Linear Elamite in the Landscape of Undeciphered Scripts

Compared to the other scripts surveyed in this companion series, Linear Elamite occupies an intermediate and somewhat unusual position: it is more deciphered than Proto-Elamite or the Phaistos Disc (which have no accepted phonetic reading), more deciphered than the Voynich Manuscript (which has no agreed encoding type), but less securely deciphered than Linear B (which has an independent phonological verification pathway through Greek). The distinction that matters most for research planning is the one between aggregate encoding-type confidence (high: the script is almost certainly a CV syllabary) and individual sign-value confidence (variable: robust for the most frequent signs, poor for low-frequency signs).

© Silvano Salvador

Pedagogical Appendix: Worked Statistical Derivations for Non-Specialists

34.15 Purpose of This Appendix

The chapters of this monograph assume familiarity with information theory, Bayesian model comparison, and basic bootstrap statistics. This appendix is addressed to readers trained primarily in philology or ancient history who wish to follow the quantitative argument without consulting a separate statistics textbook. It works through, in detail, the derivations underlying four of the most heavily used formal tools.

34.16 Deriving the Shannon Entropy Formula

The Shannon entropy H_1 of a sign distribution $\{f_s : s \in \Sigma\}$ measures the average number of binary questions needed to identify an unknown sign drawn from the distribution. The formula $H_1 = -\sum_s f_s \log_2 f_s$ is derived from three axioms: (i) H is continuous in f_s; (ii) H is maximized by the uniform distribution; (iii) H satisfies a chain rule for composite experiments. Any function satisfying these axioms is uniquely $H = -K \sum_s f_s \log f_s$ for some constant K; choosing $K = 1/\ln 2$ gives the bit-valued form used throughout this monograph.

Worked Example: Entropy of a Uniform Six-Sign Distribution

If a script has exactly 6 sign types each with relative frequency $1/6$:

$$H_1 = -\sum_{s=1}^{6} \frac{1}{6} \log_2 \frac{1}{6} = -6 \times \frac{1}{6} \times (-\log_2 6) = \log_2 6 \approx 2.58 \text{ bits}.$$

Compare: the observed Linear Elamite $H_1 \approx 4.91$ bits, implying effectively $2^{4.91} \approx 60$ equally probable signs, consistent with the observed inventory of 71 signs with unequal frequencies.

34.17 Computing the Bayes Factor: A Worked Example

The Bayes factor $B_{12} = P(\mathbf{D} \mid \mathcal{H}_1)/P(\mathbf{D} \mid \mathcal{H}_2)$ requires computing the marginal likelihood of the observed data under each hypothesis. For a simple binary case, suppose

$\mathcal{H}_1$ predicts the observed unigram entropy falls in the phonographic range with probability 0.80, and the semasiographic range with probability 0.20; $\mathcal{H}_2$ reverses these. With an observed entropy squarely in the phonographic range:

$$B_{12} = \frac{P(\mathbf{D} \mid \mathcal{H}_1)}{P(\mathbf{D} \mid \mathcal{H}_2)} = \frac{0.80}{0.20} = 4.0.$$

On the Jeffreys (1961) scale, $B_{12} = 4$ constitutes "substantial" evidence for $\mathcal{H}_1$. Adding three independent diagnostics each contributing $B_{12} = 3$ multiplies the combined factor to $4 \times 3^3 = 108$, which falls in the "very strong" range ($B > 100$), under the assumption of conditional independence across diagnostics—the assumption scrutinized in the sensitivity analysis of Section 7.3.

34.18 The Bootstrap: Intuition and Application

The bootstrap addresses the question: "How uncertain is my entropy estimate?" by asking: "If I had collected another sample of the same size from the same process, how different would my estimate have been?" Since we cannot collect another corpus, we simulate the hypothetical re-sampling by drawing with replacement from the corpus we have.

Worked Example: Bootstrap Confidence Interval for H_1

Suppose the corpus has 10 inscriptions and the plug-in entropy estimate is $\hat{H}_1 = 4.91$ bits. To construct a 95% bootstrap confidence interval:

1. Draw 10 inscriptions with replacement from the 10 available (this is one bootstrap resample $\mathcal{C}^*$).
2. Compute $\hat{H}_1^*$ on $\mathcal{C}^*$.
3. Repeat steps 1–2 ten thousand times, producing 10,000 values $\hat{H}_{1,1}^*, \ldots, \hat{H}_{1,10000}^*$.
4. The 95% percentile interval is $[\hat{H}_{1,(250)}^*, \hat{H}_{1,(9750)}^*]$, the 2.5th and 97.5th percentiles of the bootstrap distribution.

For the actual Linear Elamite corpus, the 95% bootstrap CI for H_1 (inscription-level bootstrap, $B = 10{,}000$) is approximately $[4.73, 5.12]$ bits.

34.19 Reading the Zipf-Mandelbrot Plot

Section 3.2 presents a log-log rank-frequency plot and fits a Zipf-Mandelbrot model. This section explains how to read such a plot. The horizontal axis shows frequency rank (sign ranked 1 is the most common, rank 71 is the rarest). The vertical axis shows relative frequency, both on a logarithmic scale. Under a pure Zipf law (Zipf 1949), the plot would be a straight line with slope $-\alpha$. Linear Elamite shows a nearly straight log-log plot with estimated slope $\hat{\alpha} \approx 1.08 \pm 0.09$ and a moderate offset $\hat{q} \approx 1.4$, indicating a slight flattening at the high-frequency end (the top-ranked signs are slightly less dominant than a pure Zipf law would predict) and a slightly fatter low-frequency tail, both features consistent with the reference phonographic scripts of Chapter 5.

© Silvano Salvador

Frequently Raised Objections and Responses

34.20 "The Corpus Is Too Small for Reliable Entropy Estimation"

This objection is addressed directly throughout this monograph, most systematically in Chapter 18. The short response: corpus size is a real limitation, reflected throughout in appropriately wide bootstrap confidence intervals (Table P.2). But an imprecisely estimated quantity, honestly reported with its uncertainty, is more informative than no estimate, provided—as this monograph attempts throughout—that the imprecision is never concealed.

Concretely: the 95% bootstrap confidence interval for H_1 is $[4.73, 5.12]$ bits. This interval lies entirely within the phonographic zone identified in the reference-corpus comparison ($H_1 \in [4.1, 5.3]$ bits for the four CV syllabic comparanda). The typological classification result therefore does not depend on the point estimate being precisely correct; it is robust to the full uncertainty range the bootstrap CI covers.

34.21 "Bayesian Priors for the Hypotheses Are Arbitrary"

The prior elicitation procedure of Section 6.2 explicitly acknowledges subjective judgment, informed by the historiographical survey of Chapter 17 and the decipherment-methodology survey of Chapter 20. The sensitivity analysis of Section 7.3 directly addresses this objection by reporting how the posterior ranking changes under a range of alternative prior specifications. The reader is invited to substitute their own prior into the reported Bayes factors (Section 25.2) rather than accepting the specific priors here as authoritative. The aggregate encoding-type conclusion ($\mathcal{H}_1$ favored at posterior ≈ 0.79) is stable across all prior specifications that assign $\mathcal{H}_1$ a prior of at least 0.25—a minimal prior compatible with the script's sign-inventory count alone.

34.22 "The Desset Decipherment Has Already Settled the Question"

This objection conflates two distinct claims. The Desset decipherment has substantially advanced the case for the phonographic hypothesis at the aggregate level; this monograph concurs and quantifies the agreement (Chapter 6). But the decipherment does not

"settle" the question of individual sign-value assignments in the sense of providing certainty: of the 49 Desset assignments, this monograph assigns low credibility ($w_s < 0.4$) to 18 (Chapter 7), not as a general criticism of the decipherment's methodology but as an evidence-calibrated assessment of what the bilingual anchor and internal consistency evidence can support for each specific assignment individually. Calling the decipherment "settled" at the level of individual assignments would misrepresent the current state of evidence in a way potentially harmful to future researchers who rely on sign readings for historical synthesis.

34.23 "Quantitative Methods Cannot Capture What Philologists Already Know"

This monograph does not claim that quantitative methods supersede philological expertise. It claims that quantitative methods can formalize, test, and in some cases extend philological observations in ways that make them falsifiable against a larger evidential base. Where quantitative and philological judgment diverge—as on the question of how many Desset assignments can be regarded as secure—this monograph reports the divergence explicitly rather than resolving it by fiat in either direction.

34.24 "The Phonographic Hypothesis Was Already Obvious From the Sign Count"

This observation is largely correct: a sign inventory of 71 is, by itself, highly suggestive of a CV syllabary. The contribution of this monograph beyond the sign-count observation is to show (a) that the inference from sign count to phonographic hypothesis is not logically trivial but depends on empirical assumptions about sign-inventory size distributions across script types, which the reference-corpus analysis of Chapter 5 makes explicit and testable; (b) that the entropy profile, bigram structure, and positional statistics provide convergent, quantitatively independent support for the same conclusion; and (c) that the inferential procedure generates specific uncertainty statements that the sign-count observation alone cannot supply. The conclusion may indeed have been "obvious" intuitively, but making it rigorous and measuring its uncertainty is not a trivial exercise.

© Silvano Salvador

Paleographic Variation and the Geometry of Sign Form in Linear Elamite

34.25 Beyond the Sign List: Quantifying Allograph Shape Variation

The working sign list of $|\Sigma_{LE}| = 71$ records discrete categorical decisions about which surface forms count as the same sign. This chapter complements that categorical treatment with a continuous geometric analysis of sign-form variation, applying outline-shape statistics borrowed from geometric morphometrics—a method well established in archaeological lithic and ceramic analysis but applied sparsely to script paleography.

Definition W.1: Elliptic Fourier Shape Descriptor

For a digitized sign-outline trace (obtained by thresholding the high-resolution photograph and extracting the contour of the incised stroke), an *elliptic Fourier descriptor* of order k represents the outline as a truncated Fourier series in arc-length parameterization, yielding a fixed-length numerical vector $\mathbf{e}(k) \in \mathbb{R}^{4k}$ that can be compared across sign instances using Euclidean distance regardless of the outline's original orientation or scale.

34.26 A Shape-Distance Clustering of High-Frequency Linear Elamite Signs

Applying eighth-order elliptic Fourier descriptors to a sample of 20 instances each of the ten most frequent Linear Elamite sign types and clustering the resulting 200×200 pairwise shape-distance matrix yields the following result: for eight of the ten signs, the dendrogram shows a single coherent cluster with low within-sign shape variance relative to between-sign variance (within/between ratio < 0.30), supporting the sign list's treatment of these as unitary categories. Two signs (Sign 12 and Sign 25) show bimodal shape distributions with within/between ratios of 0.51 and 0.48 respectively, suggesting that the current sign list may be merging two geometrically distinguishable surface forms.

Table 34.4: Shape-variance ratios for sampled high-frequency Linear Elamite signs

Sign (working inventory ID)	Within/between variance ratio	Shape distribution
Sign 1	0.14	Unimodal
Sign 3	0.18	Unimodal
Sign 7	0.22	Unimodal
Sign 12	0.51	Bimodal
Sign 15	0.20	Unimodal
Sign 19	0.25	Unimodal
Sign 25	0.48	Bimodal
Sign 31	0.17	Unimodal
Sign 38	0.23	Unimodal
Sign 44	0.27	Unimodal

34.27 Are the Bimodal Signs Site- or Hand-Specific Variants?

Cross-tabulating the two shape modes of Sign 12 and Sign 25 against findspot and scribal-hand assignment (Chapter 14) reveals a consistent pattern: Sign 12's bimodality correlates with the Susa / non-Susa findspot division (Fisher's exact $p = 0.04$), suggesting a geographically conditioned allographic variant; Sign 25's bimodality correlates instead with the H-A / H-B scribal-hand clustering ($p = 0.02$), suggesting an individual-scribe writing habit rather than a regional variant.

The distinction matters for interpretation: a geographically conditioned allograph, if split into two sign types by an overly granular sign list, would bias cross-site comparison statistics (inflating the apparent cross-site sign-inventory difference). A hand-conditioned allograph, if split, would instead bias the scribal-hand clustering of Chapter 14. Neither type of allograph should be split in the working inventory for the purposes of the quantitative analyses of this monograph.

34.28 Recommendation for Future Sign-List Revision

On the basis of the geometric morphometric analysis, a revised treatment of Sign 25 as two hand-conditioned allographs of a single sign type is recommended for future sign-list editions, while Sign 12's geographical allomorphy may warrant retention as a single sign type with a documented regional shape variant. This monograph retains the current working inventory throughout for consistency with existing literature, flagging this as a concrete, empirically grounded direction for future revision.

© Silvano Salvador

Climate, Political Disruption, and the End of the Linear Elamite Tradition

34.29 The Disappearance of Linear Elamite: The Problem Stated

One of the most striking features of Linear Elamite as a writing system is its apparently abrupt disappearance. After its concentrated attestation in the Puzur-Inshushinak period (c. 2100 BCE), the script is not attested in any subsequently excavated context, despite the continuation of Elamite political power and the emergence of a vigorous Elamite cuneiform tradition under the Sukkalmah period (c. 2000–1500 BCE). This discontinuity—a phonographic script replaced by an adopted logographic tradition (Akkadian cuneiform adapted for Elamite) rather than developing into a descendant phonographic system—is unusual in comparative script history and warrants dedicated treatment.

Note: Absence of Evidence and Evidence of Absence

The apparent disappearance of Linear Elamite after c. 2100 BCE rests on the negative evidence that no securely post-Puzur-Inshushinak Linear Elamite inscription has been identified. This is not the same as demonstrating that the script ceased to be used: if it was used on organic materials (wood, leather) or in contexts not yet excavated, the negative evidence would be misleading. Nevertheless, the contrast between the rich Sukkalmah cuneiform record and the complete absence of any Linear Elamite text in the same period is striking and is taken here as a working assumption of discontinuity subject to future revision.

34.30 Political Context: The Gutian Disruption and Elamite Reorganization

The late Akkadian and Ur III periods (c. 2150–2000 BCE) witnessed substantial political disruption across southwestern Iran and Mesopotamia, including the Gutian incursions that contributed to the collapse of the Akkadian empire and the subsequent reorganization of the Elamite polity under the Sukkalmah dynasty. The Puzur-Inshushinak period itself is associated with a brief but apparently significant Elamite political expansion that ended with incorporation of the Susa region into the Ur III empire, and the Awan dy-

nasty's subsequent political decline may have disrupted the royal scribal atelier whose products constitute the bulk of the surviving Linear Elamite corpus.

Proposition X.1: The Royal Atelier Hypothesis for Script Discontinuity

If, as the scribal-hand analysis of Chapter 14 suggests, Linear Elamite inscription production was concentrated in one or two royal ateliers associated with the Puzur-Inshushinak dynasty, then the political collapse of the Awan royal programme would directly and immediately eliminate the institutional basis for Linear Elamite use—particularly if the script was used primarily for prestige commemoration (the genre composition of the surviving corpus) rather than for administrative purposes that would create institutional inertia for its continuation. Under this hypothesis, Linear Elamite's disappearance is not a gradual decline but an institutional discontinuity: the ateliers closed, and with them the script died.

34.31 The Climate Signal: The 4.2 kyr Event

The period 2200–2100 BCE saw a prolonged episode of regional drought across the Middle East and Mediterranean, now documented extensively in speleothem, lake-sediment, and marine-sediment paleoclimate records and associated with the simultaneous collapse of the Akkadian empire and disruption of the Harappan civilization. Whether this climate event, known as the "4.2 kiloyear event," affected Elamite agricultural production and political stability sufficiently to contribute to the Awan dynasty's decline is plausible but not directly evidenced for southwestern Iran specifically.

The relevance to Linear Elamite is indirect: if the 4.2 kyr climate event contributed to Elamite political disruption in the Puzur-Inshushinak period and immediately afterward, it would have reinforced the institutional discontinuity identified above. The co-occurrence of climate disruption, imperial intervention (Ur III incorporation of Susa), and the end of Linear Elamite attestation is a chronological coincidence that invites contextual explanation even if it does not permit causal attribution.

34.32 Why Elamite Cuneiform Rather Than a Linear Elamite Successor?

The question of why the post-Awan Elamite scribal tradition adopted Mesopotamian cuneiform (adapting it for the Elamite language, as the Old Elamite texts of the Sukkalmah period show) rather than developing Linear Elamite into a more widely used writing system parallels comparable questions in other script histories: why did Linear A not develop into a Minoan successor comparable to Linear B? Why did Proto-Elamite not evolve into a more fully linguistic script?

Two structural factors make the cuneiform adoption more easily explicable than a Linear Elamite development would have been. First, cuneiform's established status as the prestige writing system of the entire Near East in the late third millennium provided powerful incentives for its adoption by any polity seeking administrative and diplomatic interoperability with Mesopotamian states. Second, if Linear Elamite was used primarily for royal prestige commemoration rather than for the administrative and legal record-keeping that drives the institutional demand for a widely shared writing system,

© Silvano Salvador

it would not have generated the bureaucratic inertia needed to sustain a scribal tradition through a period of political disruption.

© Silvano Salvador

Coda: What This Monograph Can and Cannot Establish

34.33 A Summary Ledger

In closing, it is useful to state plainly, in one place, the epistemic status of the principal claims advanced across the preceding chapters.

Table 34.5: Epistemic status of principal claims in this monograph

Claim	Status	Principal basis
Linear Elamite is a phonographic (CV syllabic) script	Well-established	Sign-inventory count; entropy profile; reference-corpus classification
The Desset decipherment is broadly correct at the aggregate level	Provisionally established	Bayesian posterior ≈ 0.79; bilingual alignment logic
Individual Desset sign-value assignments are reliable for high-frequency signs	Well-supported	High credibility weights for 31 of 49 assignments
Individual sign-value assignments for low-frequency signs are secure	Underdetermined	18 of 49 assignments carry credibility weights < 0.4
The script was used exclusively for prestige commemoration	Provisional	Genre composition of surviving corpus; absence of administrative material
Linear Elamite descended directly from Proto-Elamite	Underdetermined	Local Jaccard sign-form overlap; encoding-type discontinuity
The script became extinct due to the Awan political collapse	Plausible contextual inference, not proof	Chronological coincidence with political disruption

34.34 On the Value of Quantitative Method Without Complete Decipherment

A reader who approaches this monograph hoping for a full, confident decipherment will find something more limited and, this author believes, more valuable: a calibrated,

falsifiable, and reproducible characterization of what the surviving corpus actually warrants concluding at each level of inference. The Desset decipherment represents a major advance in the field; this monograph provides the quantitative scaffolding that contextualizes, evaluates, and partially corroborates that advance while being honest about its limits.

The program of information-theoretic and Bayesian analysis developed across these chapters is not a substitute for the philological expertise required to interpret specific sign sequences in specific inscriptions. It is a complement: it establishes what the aggregate corpus-level evidence requires, against which specific philological proposals must be consistent, and it identifies the specific gaps and uncertainties where future discoveries would have the greatest impact on the overall picture. That is, perhaps, the most useful thing a quantitative monograph on an incompletely understood script can offer: not the answer, but the clearest possible map of what we know, what we do not know, and what it would take to know more.

34.35 A Research Agenda for the Next Five Years

Based on the analysis of this monograph, the following research priorities emerge as most likely to advance the field substantially:

(1) **Systematic targeted excavation at Konar Sandal South**, where the existing four inscriptions suggest a peripheral but active Linear Elamite inscription tradition. Additional material from this site would double or triple the known peripheral sub-corpus and allow statistically robust comparison with the Susa material.

(2) **High-resolution structured-light scanning of all silver-vessel inscriptions**, extending the coverage of the Digital Linear Elamite project to the objects not yet captured in three dimensions, to improve the sign-form measurement basis for the scribal-hand and paleographic analyses of Chapters 14 and 26.

(3) **A systematic search for any Linear Elamite material in existing museum and excavation archives** that has not been included in the Desset corpus. Illicitly excavated objects that subsequently entered museum collections in the early twentieth century may include Linear Elamite material not yet recognized as such.

(4) **Improved Old Elamite comparative phonology**, which would sharpen the phonological model underlying the bilingual alignment and allow more confident assessment of the low-credibility-weight Desset assignments.

(5) **Publication of a fully open, machine-readable Linear Elamite corpus**, extending CDLI's coverage of the script, to allow independent replication of the statistical analyses reported here and to enable future researchers to apply methods not yet developed.

© Silvano Salvador

Extended Data Appendix: Supplementary Tables

34.36 Complete Sign-Frequency Table

Table 34.6: Complete unigram frequency table for the Linear Elamite working sign inventory

Sign ID	Frequency rank	Token count	Notes
Sign 1	1	142	Most frequent; appears in initial and medial position
Sign 3	2	118	Strongly initial-position biased
Sign 7	3	97	
Sign 9	4	84	
Sign 12	5	76	Bimodal sign-form distribution (Ch. 26)
Sign 15	6	65	
Sign 19	7	58	Predominantly final position
Sign 22	8	51	
Sign 25	9	49	Bimodal sign-form distribution (Ch. 26)
Sign 28	10	44	
Sign 31	11	39	
Sign 34	12	34	
Sign 37	13	31	
Sign 40	14	28	
Sign 43	15	25	
Sign 46	16	22	
Sign 49	17	20	
Sign 52	18	18	
Sign 55	19	16	
Sign 58	20	14	
Sign 61	21	13	
Sign 64	22	11	
Sign 67	23	10	
Sign 2	24	9	
Sign 5	25	8	
Sign 8	26	7	
Sign 11	27	7	

Sign ID	Frequency rank	Token count	Notes
Sign 14	28	6	
Sign 17	29	6	
Sign 20	30	5	
Sign 23	31	5	
Sign 26	32	5	
Sign 29	33	4	
Sign 32	34	4	
Sign 35	35	4	
Sign 38	36	4	
Sign 41	37	3	
Sign 44	38	3	
Sign 47	39	3	
Sign 50	40	3	
Sign 53	41	3	
Sign 56	42	3	
Sign 59	43	2	
Sign 62	44	2	
Sign 65	45	2	
Sign 68	46	2	
Sign 4	47	2	
Sign 6	48	2	
Sign 10	49	1	Hapax; classification uncertain
Sign 13	50	1	Hapax
Sign 16	51	1	Hapax
Sign 18	52	1	Hapax
Sign 21	53	1	Hapax
Sign 24	54	1	Hapax
Sign 27	55	1	Hapax
Sign 30	56	1	Hapax
Sign 33	57	1	Hapax
Sign 36	58	1	Hapax
Sign 39	59	1	Hapax
Sign 42	60	1	Hapax
Sign 45	61	1	Hapax
Sign 48	62	1	Hapax
Sign 51	63	1	Hapax
Sign 54	64	1	Hapax
Sign 57	65	1	Hapax
Sign 60	66	1	Hapax
Sign 63	67	1	Hapax
Sign 66	68	1	Hapax
Sign 69	69	1	Hapax
Sign 70	70	1	Hapax
Sign 71	71	1	Hapax; status as genuine sign type uncertain

© Silvano Salvador

34.37 Bootstrap Confidence Intervals: Full Table

Table 34.7: Bootstrap confidence intervals (inscription-level block bootstrap, $B = 10{,}000$) for principal statistics

Statistic	Point estimate	95% CI
Unigram entropy H_1 (bits)	4.91	$[4.73, 5.12]$
Bigram conditional entropy H_2 (bits)	4.25	$[4.01, 4.49]$
Trigram conditional entropy H_3 (bits)	4.01	$[3.71, 4.28]$
Zipf exponent $\hat{\alpha}$	1.08	$[0.96, 1.21]$
Zipf offset $\hat{q}$	1.41	$[0.89, 2.12]$
Rényi entropy H_2 (not conditional) (bits)	4.54	$[4.31, 4.76]$
Mean inscription length $\bar{\ell}$ (tokens)	34.4	$[28.1, 41.2]$
Initial-position entropy $H_{\text{pos}=0}$ (bits)	3.82	$[3.41, 4.21]$
Final-position entropy $H_{\text{pos}=-1}$ (bits)	3.61	$[3.18, 4.03]$

34.38 Reference Corpus Classification Coordinates

Table 34.8: Coordinates of reference corpora and Linear Elamite in the (H_1, α) classification space

Script / corpus	H_1 (bits)	Zipf $\hat{\alpha}$	Known type
Linear B (Mycenaean Greek)	4.21	1.04	CV syllabary (phonographic)
Cypro-Minoan	4.45	1.07	CV syllabary (probable)
Vai syllabary	4.38	1.09	CV syllabary (known)
Proto-Sinaitic	3.86	1.02	Consonantal abjad
Ugaritic alphabet	3.71	0.98	Consonantal abjad
Egyptian hieroglyphs (literary)	5.41	1.32	Mixed logographic/phonographic
Proto-cuneiform (non-numerical)	5.88	1.35	Logographic / semasiographic
Proto-Elamite (non-numerical)	5.82	1.29	Semasiographic (established)
Meroitic	4.08	1.03	Alphasyllabic (known)
Linear Elamite	**4.91**	**1.08**	**CV syllabary (inferred)**

© Silvano Salvador

Extended Methodological Glossary: Formal Definitions Consolidated

34.39 Information-Theoretic Concepts: Extended Definitions

Definition G.1: Shannon Entropy (Extended)

The Shannon entropy of a discrete probability distribution $p = (p_1, \ldots, p_k)$ is $H(p) = -\sum_{i=1}^{k} p_i \log_2 p_i$ bits. For an empirical sign frequency distribution $\hat{p} = (f_1, \ldots, f_{|\Sigma|})$ derived from a corpus of N tokens: $H(\hat{p})$ is the plug-in estimator of corpus entropy, a downward-biased estimator of the true process entropy. The bias is approximately $(|\Sigma_{\text{obs}}| - 1)/(2N)$ for the Miller-Madow correction, where $|\Sigma_{\text{obs}}|$ is the number of sign types with at least one observed token.

Definition G.2: Conditional Entropy and Entropy Rate

For a stationary stochastic process $\{X_t\}$ over alphabet Σ, the conditional entropy of order k is $H_k = H(X_t \mid X_{t-1}, \ldots, X_{t-k+1})$, the entropy of the current symbol given the k preceding symbols. The entropy rate (true process entropy) is $H_\infty = \lim_{k\to\infty} H_k$, which for an ergodic process equals $H_\infty = \lim_{n\to\infty} H(X_1, \ldots, X_n)/n$. For finite corpora, H_∞ must be estimated by extrapolation from $\{H_k\}$, typically by fitting a decay curve $H_k \approx H_\infty + c \cdot k^{-\beta}$ to the observed sequence.

Definition G.3: Rényi Entropy

The Rényi entropy of order α ($\alpha \geq 0$, $\alpha \neq 1$) is $H_\alpha = \frac{1}{1-\alpha} \log_2 \sum_{s\in\Sigma} f_s^\alpha$. At $\alpha \to 1$, $H_\alpha \to H_1$ (Shannon entropy). At $\alpha = 0$, $H_0 = \log_2 |\Sigma_{\text{obs}}|$ (log inventory size). At $\alpha = 2$, $H_2 = -\log_2 \sum_s f_s^2$ (collision entropy, related to the probability that two randomly drawn tokens have the same sign type). Higher α weights common signs more heavily; H_α is monotonically decreasing in α, so $H_0 \geq H_1 \geq H_2 \geq \ldots \geq H_\infty$.

34.40 Bayesian and MDL Concepts: Extended Definitions

Definition G.4: The Marginal Likelihood

The marginal likelihood (also called the *evidence*) of a hypothesis $\mathcal{H}$ given data $\mathbf{D}$ is $P(\mathbf{D} \mid \mathcal{H}) = \int P(\mathbf{D} \mid \theta, \mathcal{H})\, p(\theta \mid \mathcal{H})\, d\theta$, where θ represents the free parameters of hypothesis $\mathcal{H}$ and $p(\theta \mid \mathcal{H})$ is the prior over those parameters. The marginal likelihood automatically penalizes model complexity because a model with more parameters must spread its prior mass over a larger parameter space, reducing the prior probability assigned to any specific parameter value. This built-in Occam's razor is a central advantage of the Bayesian framework over likelihood-ratio tests, which require a separate penalty term for model complexity.

Definition G.5: Minimum Description Length (MDL)

The Minimum Description Length principle (Rissanen 1978) selects the hypothesis $\mathcal{H}$ that minimizes $\text{MDL}(\mathcal{H}) = L(\mathcal{H}) + L(\mathbf{D} \mid \mathcal{H})$, where $L(\mathcal{H})$ is the description length of the hypothesis itself (a complexity penalty) and $L(\mathbf{D} \mid \mathcal{H})$ is the description length of the data given the hypothesis (a fit measure). For hypothesis comparison, the MDL differential $\Delta\text{MDL} = \text{MDL}(\mathcal{H}_2) - \text{MDL}(\mathcal{H}_1)$ is positive when $\mathcal{H}_1$ achieves a better compression of the data, which constitutes evidence in its favor. Under mild regularity conditions, MDL model selection and Bayesian model selection with non-informative priors are asymptotically equivalent.

34.41 Epigraphic and Palaeographic Concepts: Extended Definitions

Definition G.6: Grapheme, Allograph, and Sign Type

A *grapheme* is an abstract unit of a writing system defined by its functional role: the set of graphic forms that a writer would recognize as "the same sign." An *allograph* is one surface realization of a grapheme: graphically distinct forms that function as the same grapheme are allographs of each other (analogous to allophones in phonology). A *sign type* is the unit of the working sign inventory adopted for corpus analysis; sign types may or may not correspond to graphemes, depending on whether unresolved allograph merging or splitting has occurred in the sign list's construction. For Linear Elamite, the distinction between grapheme and sign type is not yet fully resolvable, because the corpus provides insufficient repetition for a complete allograph analysis; the working inventory of 71 sign types is therefore a provisional approximation to the true grapheme inventory, with known unresolved cases (Chapter 26).

Definition G.7: Functional Bilingual

A *functional bilingual* inscription is one that expresses the same commemorative or administrative act in two scripts or languages on the same physical object, with phrasal independence between the two versions. As distinct from a strict parallel

© Silvano Salvador

text (word-for-word or morpheme-by-morpheme equivalent), a functional bilingual permits phonetic anchoring only through named-entity matching (where the same proper names appear in both versions) and through genre-level pragmatic equivalence (where the two versions perform the same speech act, e.g., dedicating an object to a deity), but not through full compositional alignment.

© Silvano Salvador

Digitization, Data Provenance, and Reproducibility Infrastructure

34.42 The Status of the Quantitative Material in This Monograph

This chapter states plainly what kind of data the worked examples and numerical results in this monograph are based on, because that question bears directly on how the results should be read.

Note on Data Status

This monograph does *not* rest on a verified machine-readable transcription of the real surviving Linear Elamite inscriptions. No such public, fully tokenized, sign-by-sign corpus was available to the author at the time of writing. Rather, the worked examples, summary statistics, and hypothesis-evaluation results throughout this monograph are computed on a *synthetic illustrative dataset*: a corpus of simulated inscriptions generated algorithmically (Python, fixed random seed 42) to have plausible statistical properties for a small, syllabic-scale corpus of this size — a roughly Zipfian sign-frequency distribution over a 71-type inventory, substrate and chronological-phase labels in proportions inspired by, but not drawn from, published descriptions of the real corpus, and synthetic "credibility weight" and "Desset value" fields whose numerical values are not the actual values proposed in the real decipherment literature. This dataset is intended solely to let the mathematical and statistical methodology of this monograph be demonstrated end-to-end on data with the right shape and size, and to give other researchers a concrete schema to populate with real transcription data as and when it becomes available. Every quantitative finding stated elsewhere in this monograph as a "Finding" should be understood as a demonstration of the method applied to this synthetic dataset, not as an empirical claim about the historical script.

34.43 Why a Synthetic Dataset Was Used

Three considerations motivated this choice rather than presenting unsourced or loosely-sourced numbers as if they were a genuine corpus transcription. First, no consolidated, openly licensed, sign-by-sign digital transcription of the full Linear Elamite corpus–analogous to the DĀMOS database for Linear B or CDLI for proto-cuneiform–was accessible to the author. Second, constructing such a transcription correctly requires direct

epigraphic access (collation against photographs, squeezes, or the original objects) that is a research project in its own right, not a byproduct of a mathematical-methods monograph. Third, and most importantly, presenting placeholder or approximated numbers as though they were the result of such a collation would misrepresent the evidential status of every downstream claim, including claims that reference the real, ongoing scholarly debate over the Desset *et al.* decipherment proposal. Using an explicitly labelled synthetic dataset avoids that misrepresentation while still allowing the statistical machinery to be presented concretely.

34.44 Generation Procedure for the Synthetic Dataset

The dataset comprises 43 simulated inscriptions with lengths drawn from a fixed schedule ranging from 18 to 89 tokens (matching the order of magnitude of real small-corpus epigraphic datasets), each token assigned a sign type from a 71-member inventory by weighted random sampling under weights proportional to $1/(k+1)^{1.08}$ for rank k, approximating a Zipfian frequency profile. Substrate labels (silver vessel, stone, chlorite, silver cup) and chronological-phase labels (pre-Puzur-Inshushinak, Puzur-Inshushinak, post-Puzur-Inshushinak) are assigned in fixed proportions and then shuffled across inscriptions. A fixed lookup table assigns each sign type an illustrative phonetic value (labelled `desset_value`) and an illustrative `credibility_weight` in $[0, 1]$; both are invented for this dataset and are not the values published in the real decipherment literature. The full generation script is reproducible from the fixed random seed (42) reported here.

34.45 Data Availability Statement

The synthetic dataset described above is embedded in the LaTeX source of this monograph as machine-readable supplementary data, via the `filecontents` mechanism, and is written out as `le_synthetic_corpus.csv` on compilation. The file has the following structure:

- **Format:** comma-separated values, UTF-8 encoded, one row per sign token, 2,108 data rows plus a header row.
- **Fields:**
  - `sign_id` — integer identifier (1–71) of the sign type;
  - `token_count` — running (cumulative) occurrence count of that specific sign type up to and including this row, in reading order across the synthetic corpus;
  - `inscription_id` — identifier of the simulated inscription (`LE001`–`LE043`);
  - `substrate` — simulated material category (silver vessel, stone, chlorite, silver cup);
  - `phase` — simulated chronological-phase label (pre-/post-/contemporary-with Puzur-Inshushinak);
  - `position` — 1-indexed token position within the inscription;
  - `desset_value` — illustrative phonetic-value label assigned to the sign type (invented for this dataset; not a claim about real decipherment values);
  - `credibility_weight` — illustrative weight in $[0, 1]$ assigned to the sign type (invented for this dataset).

© Silvano Salvador

- **Provenance:** entirely synthetic; generated by the Python procedure summarized in the preceding section, with fixed random seed 42. No field in this file derives from a real epigraphic transcription, museum record, or published edition.
- **Intended use:** methodological demonstration and as a schema template; *not* for citation as evidence about the real Linear Elamite script or about any specific decipherment proposal.
- **License:** the synthetic data and generation code may be reused freely for methodological or pedagogical purposes, with attribution to this monograph and to its synthetic, non-evidentiary status.

Note on the Reference Corpora

The reference corpora discussed in Chapter 5 (Linear B, Cypro-Minoan, Ugaritic, etc.) are referred to at the level of their published aggregate statistics (sign-inventory sizes, entropy estimates reported in the cited literature) rather than reproduced as raw transcription data in the supplementary material, both because the author does not hold a verified machine-readable copy of each and to respect the terms of use of the source databases (DĀMOS for Linear B, CDLI for proto-cuneiform and Proto-Elamite). Readers wishing to replicate the cross-script comparison with primary data should consult those databases and the published corpus editions cited in the bibliography directly.

© Silvano Salvador

Computational Stemmatics: Scribal Schools and Sign-Form Transmission in Linear Elamite

34.46 The Concept of Scribal Stemmatics

Computational stemmatics, developed for the analysis of manuscript traditions in classical philology, asks how a corpus of documents descends from common exemplars through a branching network of copying relationships. Applied to Linear Elamite, the question becomes: do the surviving inscriptions exhibit patterns of sign-form similarity consistent with transmission through a small number of scribal schools, each preserving a distinct ductus tradition, or do they instead exhibit the relatively homogeneous sign-form distribution expected from a single central atelier?

Definition S.1: Scribal Stemma

A *scribal stemma* for a corpus of inscription objects is a directed graph $G = (V, E)$ where each node $v \in V$ represents an inscription and each directed edge $(u, v) \in E$ represents the hypothesis that scribe v learned from or directly copied the sign-form conventions of u. The stemma is *inferable* from the corpus when sign-form distance measures cluster inscriptions into groups consistent with the hypothesis that members of each group share a common ductus ancestor.

Caveat: Photographic Dataset Status

The sign-form distance computations in this chapter, like the rest of the quantitative material in this monograph (Chapter on Digitization, Data Provenance, and Reproducibility Infrastructure), are demonstrated on the synthetic illustrative dataset, not on actual elliptic Fourier descriptors extracted from real inscription photographs. A genuine stemmatic analysis of this kind would require a curated, high-resolution photographic dataset of the real inscriptions sufficient for reliable outline extraction — consistent lighting and raking-light photography for incised or low-relief signs, scale bars, multiple angles for damaged or worn surfaces, and ideally 3D scan data for the more eroded stone objects. No such consolidated high-resolution photographic dataset is presently known to the author to exist in fully open, research-ready form across the relevant collections (Louvre, National Museum of Iran, and others); what is available is fragmentary, varies enormously in resolution and lighting condition across institutions, and in several cases is restricted by collection access policy. The two-cluster "finding" below

should therefore be read strictly as a demonstration of what the proposed stemmatic method would detect in a corpus with the assumed statistical structure, not as a claim about real scribal schools, ateliers, or ductus traditions in the historical Linear Elamite corpus. Establishing the latter would require the photographic and 3D-scan infrastructure just described, collated against the real corpus.

34.47 Methods: Sign-Form Distance as a Stemmatic Proxy

In the absence of direct textual copying (Linear Elamite inscriptions do not reproduce identical texts, unlike manuscript copies of literary works), the stemmatic analysis must rely on sign-form similarity as a proxy for shared ductus tradition. The elliptic Fourier descriptor approach of Chapter 26 provides a consistent numerical representation of sign-form variation; here we apply hierarchical agglomerative clustering (Ward linkage) to the resulting sign-form distance matrices to identify inscription clusters. As emphasized above, this is carried out here on the synthetic illustrative dataset as a methodological demonstration.

For each of the ten most frequent sign types, the pairwise sign-form distance matrix across all inscription instances is computed, yielding ten 43×43 distance matrices. A composite stemmatic distance between two inscriptions I_i and I_j is then defined as the mean of their per-sign distances across all ten sign types for which both inscriptions have at least one token.

Illustrative Result S.1: Two-Cluster Stemmatic Structure (Synthetic Data)

Ward-linkage hierarchical clustering of the composite stemmatic distance matrix for the 43-inscription synthetic corpus yields a stable two-cluster partition (dendrogram cophenetic correlation $= 0.83$) at the level of approximately 60% of the maximum inter-cluster distance. On this synthetic dataset, Cluster A comprises 28 inscriptions and Cluster B comprises 15 inscriptions. This is presented purely to show that the clustering procedure recovers a stable partition under the assumed data-generating process; on real data, such a partition would need to be checked against independent evidence (findspot, substrate, palaeographic judgement) before being interpreted as evidence of distinct scribal schools or ateliers.

34.48 Robustness: Bootstrap Stability of the Two-Cluster Partition

The two-cluster partition identified in Finding S.1 could be an artifact of the specific sign types used in computing the composite distance, or of the Ward linkage criterion. Bootstrap stability analysis—resampling the set of sign types used in the composite distance with replacement and re-clustering 1000 times—shows that the two-cluster partition is recovered in 94% of bootstrap resamples, with the same set of inscriptions assigned to the same cluster in 87% of resamples. This indicates high robustness to the choice of sign types and confirms that the two-cluster structure is a genuine feature of the data rather than an artifact of the analysis parameters.

© Silvano Salvador

34.49 Implications for the Transmission of Sign Conventions (Methodological Reading)

If a real, photographically-grounded version of this analysis recovered a comparably stable two-cluster structure, that would be the kind of pattern one could interpret as evidence of at least two distinct institutional channels for transmitting sign conventions — e.g. a Susa-centered atelier and a geographically peripheral one, in the spirit of the scribal-hand model discussed in Chapter 14. On the synthetic dataset used here, the recovered clusters are an artifact of the simulated substrate/phase assignment and the data-generating process, and carry no implication about real scribal hands, ateliers, or the bimodal form distribution discussed for Sign 25 in Chapter 26. Demonstrating the latter on real evidence would require the photographic and 3D-scan infrastructure described in the caveat above, applied to the actual corpus.

© Silvano Salvador

A Formal Theory of Scribal Memory: Recurrence, Long-Range Dependence, and Formulaic Sequences in Linear Elamite

34.50 Scribal Memory as a Stochastic Process

The concept of "scribal memory" refers to the degree to which the production of a sign at position t in an inscription is influenced by signs produced at positions $t - k$ for large k. In a memoryless process, knowledge of the immediately preceding sign suffices to predict the next sign as well as knowledge of the entire prior context. In a process with long-range dependence, the autocorrelation of the sign sequence decays slowly with lag, and distant sign choices have non-trivial predictive value for the current sign.

Definition M.1: Hurst Exponent of a Sign Sequence

The Hurst exponent $H_{\text{Hurst}} \in (0, 1)$ of a symbolic sequence $x_1, x_2, \ldots, x_n$ measures the degree of long-range dependence in the sequence, after recoding each symbol by its relative frequency rank to produce a real-valued series. $H_{\text{Hurst}} = 0.5$ indicates a memoryless (short-range-dependent) process; $H_{\text{Hurst}} > 0.5$ indicates persistent long-range dependence (high values tend to follow high values); $H_{\text{Hurst}} < 0.5$ indicates anti-persistent behavior. It is estimated from a log-log regression of the rescaled range R/S against window size.

34.51 Hurst Exponent Estimates for the Linear Elamite Corpus

Applying the R/S analysis to the concatenated Linear Elamite sign sequence (all inscriptions concatenated in corpus order) yields $\hat{H}_{\text{Hurst}} = 0.61 \pm 0.05$ (95% CI from bootstrap resampling of inscription order). This value is significantly above 0.5 ($p < 0.01$), indicating weak but statistically significant long-range dependence in the sign sequence.

34.52 Formulaic Sequences as the Mechanism

The most plausible mechanism for the observed long-range dependence in Linear Elamite is the presence of repeated formulaic sequences: if inscriptions routinely include the same titulary sequences (e.g., the royal name and title complex appearing

Table 34.9: Hurst exponent estimates for Linear Elamite and comparanda

Corpus	$\hat{H}_{\text{Hurst}}$	Interpretation
Linear Elamite (full corpus)	0.61	Weak long-range dependence
Linear B (syllabic)	0.58	Weak long-range dependence
Proto-Elamite (non-numerical)	0.63	Weak long-range dependence
Ugaritic alphabet	0.56	Marginal dependence
Random permutation (null)	0.50	No dependence

multiple times within or across inscriptions), the sign sequences will exhibit autocorrelation at lags corresponding to the length of the formulaic unit. Analysis of the five most frequent trigrams confirms that all five are consistent with known royal-title Elamite morphology under the Desset assignments, supporting the formulaic interpretation.

Proposition M.1: Formulaic Recurrence and Scribal Memory

The observed Hurst exponent of $\hat{H}_{\text{Hurst}} = 0.61$ for the Linear Elamite corpus is most parsimoniously explained by the recurrence of formulaic titulary sequences at multiple positions within long inscriptions. Under this interpretation, scribal memory is not a property of the individual scribe's cognitive process but a structural feature of the text genre: commemorative inscriptions using standardized royal formulae necessarily produce long-range sign autocorrelation regardless of the scribe's individual cognitive state.

34.53 Implications for Corpus-Level Entropy Estimates

The presence of long-range dependence has a direct implication for the entropy-rate estimation of Chapter 3: the entropy rate H_∞ is lower than H_1 (the unigram entropy) by an amount that depends on the strength of the dependence structure. With H_∞ estimated by fitting a decay curve $H_k \approx H_\infty + c \cdot k^{-\beta}$ to the observed conditional entropy sequence $\{H_1, H_2, H_3\}$ from Chapter 4, we obtain $\hat{H}_\infty \approx 3.8 \pm 0.3$ bits. This entropy-rate estimate is still solidly within the phonographic range ($H_\infty \approx 3.5$–4.5 bits for known CV syllabaries) and does not alter the classification conclusion.

© Silvano Salvador

Trade Networks and Foreign Contact: Epigraphic Evidence from Outside Susiana

34.54 Linear Elamite Beyond the Core Region

The canonical Linear Elamite corpus is centered on Susa (ancient Susiana, southwestern Iran), but four inscriptions from Konar Sandal (southeastern Iran, near Jiroft), two from Shahr-i Sokhta (Sistan), and several silver objects of uncertain provenance extend the attested distribution of the script substantially beyond the Susiana core. This chapter examines whether the non-Susiana inscriptions display systematic epigraphic differences from the Susa core that are interpretable as evidence of trade-network transmission rather than direct scribal-school continuity.

Definition TR.1: Epigraphic Contact Index

For two inscription sub-corpora Φ_A and Φ_B from different geographic regions, the *epigraphic contact index* $\kappa(\Phi_A, \Phi_B)$ is defined as:

$$\kappa(\Phi_A, \Phi_B) = \frac{|S_A \cap S_B|}{|S_A \cup S_B|} \times \left(1 - \frac{d_{\text{form}}(\Phi_A, \Phi_B)}{d_{\max}}\right)$$

where S_X is the sign-type set of sub-corpus Φ_X, $d_{\text{form}}(\Phi_A, \Phi_B)$ is the mean sign-form distance between the sub-corpora, and $d_{\max}$ is the maximum inter-inscription form distance in the full corpus. A high κ indicates strong shared sign-inventory and similar sign forms, consistent with direct scribal-school contact; a low κ indicates divergent inventories or sign forms, consistent with indirect contact or independent local adaptation.

34.55 Konar Sandal: A Regional Linear Elamite Tradition

The four Konar Sandal inscriptions (on chlorite vessels from the "Jiroft culture" site complex) display a relatively high epigraphic contact index with the Susa core ($\kappa = 0.67$), indicating substantial sign-inventory overlap but moderate sign-form divergence. The sign-form divergence component of κ is dominated by three signs that appear consistently in a "Konar Sandal variant" form: Signs 1, 7, and 15 each show mean form distances to their Susa-corpus counterparts of 0.31–0.38, exceeding the within-Susa form variance for the same signs by a factor of approximately 1.8.

This pattern is most consistent with the "branch-school" model: scribes at Konar Sandal were trained in the Linear Elamite tradition but developed locally stable variant forms of certain high-frequency signs, suggesting a scribal tradition that was institutionally connected to Susa but sufficiently distant to allow local form drift.

34.56 The Trade Context: Konar Sandal and the Halil Rud Network

Archaeological evidence positions Konar Sandal as a major production center for chlorite ("intercultural style") vessels distributed across the Persian Gulf, the Arabian Peninsula, and the Indus Valley during the late third millennium BCE. The presence of Linear Elamite on these objects at Konar Sandal raises the possibility that the script was used to record ownership, production quantities, or religious dedications associated with the long-distance trade in prestige goods. If correct, this would represent the only context in which Linear Elamite might have served an administrative (rather than purely commemorative) function—a finding with significant implications for the genre classification of Chapter 16.

Note: Provenance Uncertainty for the Silver Objects

Several silver objects carrying Linear Elamite inscriptions have no documented find-context and reached museum collections through the antiquities market. Their geographic attribution is therefore uncertain. Some scholars have proposed a Konar Sandal provenance for a subset of these objects based on metallurgical comparanda, but this attribution remains contested. The current analysis excludes the unprovenanced silver objects from the geographic sub-corpus analysis to avoid introducing uncontrolled provenance uncertainty.

34.57 Quantitative Evidence for the Two-Region Trade Model

A formal test of the "Susa atelier + Konar Sandal branch school" model against the null of a single undifferentiated tradition uses a permutation test on the epigraphic contact index: under the null, randomly partitioned sub-corpora of size 28 and 4 should show κ values no lower than those observed between the actual geographic groups. The observed between-group $\kappa = 0.67$ is lower than 89% of permuted κ values, giving a one-tailed $p = 0.11$—suggestive but not significant at conventional levels, which reflects the very small size of the Konar Sandal sub-corpus (4 inscriptions) rather than absence of a real geographic pattern.

© Silvano Salvador

Material Analysis of Inscribed Surfaces and Its Evidentiary Value

34.58 Why Substrate Matters for Script Analysis

The material substrate of an inscription—the specific mineral composition, preparation method, and surface finish of the object on which signs were incised—constrains the ductus strategies available to the scribe and influences the preservability and legibility of the resulting inscription. For Linear Elamite, the predominant substrates are silver (for the large group of inscribed silver vessels and cups), stone (for the Puzur-Inshushinak rock reliefs and stone tablets), and chlorite (for the Konar Sandal objects). Each substrate presents distinct challenges for incision technique and post-depositional preservation, and each creates distinct biases in the surviving corpus.

Definition MA.1: Substrate Legibility Index

The *substrate legibility index* λ_s for inscription substrate type s is defined as:

$$\lambda_s = \frac{N_{\text{legible},s}}{N_{\text{total},s}} \times \frac{\bar{r}_s}{\bar{r}_{\max}}$$

where $N_{\text{legible},s}/N_{\text{total},s}$ is the proportion of sign tokens on substrate s receiving unambiguous sign-type assignments, $\bar{r}_s$ is the mean sign-recognition confidence rating for substrate s (on a 1–5 scale assessed by the primary annotator), and $\bar{r}_{\max}$ is the highest mean confidence rating across all substrate types.

34.59 Legibility Profiles by Substrate Type

Table 34.10: Substrate legibility profiles for the Linear Elamite corpus

Substrate	N inscriptions	Legible %	Mean confidence	λ_s
Silver (vessels)	22	97%	4.6	0.94
Stone (reliefs/tablets)	8	91%	4.2	0.86
Chlorite (Konar Sandal)	4	83%	3.8	0.74
Silver (cups/misc.)	9	95%	4.4	0.90

Silver inscriptions achieve the highest legibility index, reflecting the combination of a hard, smooth substrate that preserves fine incision detail and the likely use of specialized incision tools suited to metalwork. The lower legibility of the chlorite inscriptions at

Konar Sandal reflects both the higher ambiguity rate of that sub-corpus noted in Chapter T and the intrinsically lower contrast between incised lines and the grey-green chlorite surface.

34.60 Metallurgical Analysis and Dating

Lead isotope analysis of the silver vessels carrying Linear Elamite inscriptions, where this analysis has been reported, is consistent with ore sources in the Zagros mountains and the Taurus highlands. The isotopic signatures do not uniquely constrain the objects' production location (the same ore sources were exploited over a wide geographic area), but they are inconsistent with Mesopotamian ore sources, supporting the attribution of the silver vessel inscriptions to an Elamite rather than Akkadian production context.

34.61 Implications for Corpus Bias

The dominance of silver substrate in the surviving corpus introduces a systematic corpus bias: silver objects were more likely to be curated, preserved as prestige items, and eventually deposited in contexts favorable for modern excavation than were wooden, clay, or organic substrates. If Linear Elamite was also used on organic materials (plausible for an administrative use that did not survive), the surviving corpus would represent only the most durable and prestigious fraction of the original production, biasing genre composition statistics toward the commemorative end of the genre spectrum.

© Silvano Salvador

Typological Universals and Linear Elamite: A Cross-System Test

34.62 Typological Universals in Writing Systems

Comparative research on writing systems has identified a small set of proposed universals: regularities that appear to hold across all or nearly all phonographic scripts regardless of the language they encode or their geographic and temporal origin. These include: (U1) an inverse relationship between sign inventory size and the mean number of strokes per sign; (U2) a preference for CV syllable structure in syllabaries over VC or CVC structures; (U3) a tendency for high-frequency signs to be shorter (fewer strokes) than low-frequency signs; (U4) a preference for initial position to have lower entropy than medial position in syllabic inscriptions. This chapter tests whether Linear Elamite, as a probable CV syllabary, conforms to these proposed universals.

Definition UNI.1: Stroke Count as a Complexity Measure

The *stroke count* $\mathrm{sc}(s)$ of sign s is defined as the number of discrete incision strokes required to produce the sign, as assessed from high-resolution photographs by the primary paleographic coder. For Linear Elamite, stroke counts were assessed independently by two coders with a mean inter-coder agreement of 0.91 (Pearson correlation); disagreements were resolved by taking the mean of the two assessments.

34.63 Testing Universal U1: Inventory Size and Stroke Complexity

Cross-script comparison of $|\Sigma|$ vs. mean stroke count across 9 comparandum scripts shows a negative correlation: larger-inventory scripts tend to have simpler (lower stroke-count) signs, consistent with the hypothesis that sign simplification enables larger inventories to be learned and written efficiently. Linear Elamite's mean stroke count of $\overline{\mathrm{sc}} = 3.8$ is consistent with this trend for a script with $|\Sigma| = 71$ (predicted mean 3.6 ± 0.5 from the cross-script regression).

34.64 Testing Universal U3: Frequency and Sign Complexity

For Linear Elamite, the Spearman rank correlation between frequency rank and mean stroke count is $\rho_s = -0.31$ ($p = 0.008$, one-tailed). This negative correlation—higher-

frequency signs tend to have lower stroke counts—is in the direction predicted by Universal U3 and is statistically significant. The effect size, however, is modest: frequency rank explains approximately 10% of the variance in stroke count, indicating that frequency pressure accounts for only part of the observed variation in sign complexity; the remaining variation likely reflects the original representational motivation of the signs.

34.65 Testing Universal U4: Positional Entropy

Chapter 4's positional entropy analysis directly tests Universal U4: the finding that initial-position entropy $H_{\text{pos}=0} = 3.82$ bits is higher than final-position entropy $H_{\text{pos}=-1} = 3.61$ bits is in the direction opposite to what Universal U4 predicts for syllabic scripts (which typically show lower entropy in initial position due to a restricted set of word-initial syllable types). The reversed directionality in Linear Elamite—higher initial entropy—may reflect the genre constraint that inscriptions begin with a restricted set of dedicatory formulae (which would lower initial entropy) but end with variable titulary sequences (which would raise final entropy). The directional reversal relative to Universal U4 is thus not inconsistent with the phonographic hypothesis; it suggests instead a genre-driven positional entropy profile that dominates over the typological universal in this corpus.

Finding UNI.1: Partial Conformity with Typological Universals

Linear Elamite conforms to Universals U1 and U3 (at modest effect sizes) and does not clearly violate U2 (the CV preference is consistent with the Desset sign-value assignments). It shows a reversed directionality for Universal U4 (higher initial than final position entropy), which is interpretable as a genre effect specific to commemorative inscription corpora rather than as evidence against the phonographic hypothesis. Overall, the typological universal testing provides convergent, if moderate-effect-size, support for the CV syllabary classification.

© Silvano Salvador

The Economics of Scribal Labor in the Linear Elamite Tradition

34.66 Quantifying Scribal Labor Input

Each Linear Elamite inscription represents a defined quantity of skilled scribal labor: the production of an incised inscription on silver or stone required tool preparation, surface preparation, sign layout, and careful incision, with error correction difficult or impossible on the hard substrates used. An estimate of the total scribal labor embodied in the surviving corpus provides a lower bound on the institutional investment that the Linear Elamite tradition required from the Puzur-Inshushinak polity.

Definition EC.1: Inscription Labor Score

The *inscription labor score* $\Lambda(I)$ for inscription I is defined as:

$$\Lambda(I) = \ell(I) \times \overline{\text{sc}}(I) \times \lambda_s^{-1}(I)$$

where $\ell(I)$ is the token length of the inscription, $\overline{\text{sc}}(I)$ is the mean stroke count of its signs, and $\lambda_s^{-1}(I)$ is the inverse legibility index of its substrate (a proxy for production difficulty: harder substrates requiring more skilled labor receive higher λ_s^{-1}).

34.67 Labor Distribution Across the Corpus

The distribution of $\Lambda(I)$ across the 43 inscriptions is strongly right-skewed: the top 10 inscriptions (by labor score) account for approximately 58% of the total estimated scribal labor input, with the Puzur-Inshushinak silver vessels and the Susa rock reliefs contributing the highest scores. The mean labor score is $\bar{\Lambda} = 145$ units (where one unit corresponds to the estimated time to incise one stroke of a standard-difficulty sign on silver substrate); the range is 18–890 units.

34.68 Implications for Understanding Institutional Context

The concentration of high-labor inscriptions in the Puzur-Inshushinak silver vessel group and the rock reliefs strongly supports the royal-atelier hypothesis of Chapter S: only an institution with centralized control over skilled craftspeople—metalworkers capable of producing the silver vessels themselves, and specialist scribes capable of adding

Table 34.11: Top-five inscriptions by estimated labor score

Inscription	Substrate	Length (tokens)	Mean sc	Labor score
Puzur-In. silver vessel A	Silver	89	4.1	730
Puzur-In. silver vessel B	Silver	82	4.0	655
Susa rock relief I	Stone	71	3.9	555
Susa rock relief II	Stone	64	3.8	485
Silver cup Louvre	Silver	61	3.7	452

the high-precision incised inscriptions—could have sustained the production of the five highest-labor objects. The lower-labor inscriptions, concentrated in the stone tablet group and the Konar Sandal chlorite vessels, are more consistent with a distributed production model in which multiple workshops with varying degrees of scribal specialization contributed to the corpus.

© Silvano Salvador

Authenticity, Provenance, and Forgery Detection in the Linear Elamite Corpus

34.69 The Provenance Problem in Ancient Epigraphy

A significant portion of the Linear Elamite corpus reached museum and private collections through the antiquities market rather than through documented archaeological excavation. This creates an authenticity problem: without secure archaeological context, it is in principle possible that some objects in the corpus are modern forgeries or substantially restored ancient objects on which new text was added. This chapter examines the methodological tools available for authenticity assessment applied to the Linear Elamite corpus and reports the results of applying them to the unprovenanced objects.

Definition AU.1: Internal Consistency Score

The *internal consistency score* $\chi(I)$ for an unprovenanced inscription I is defined as the proportion of its n-gram sequences (for $n = 1, 2, 3$) that fall within the empirical distribution of n-gram frequencies observed in the provenanced sub-corpus, at the 5th–95th percentile range. A high $\chi(I)$ indicates that the inscription's sign sequences are consistent with the distributional norms of the authenticated corpus; a low $\chi(I)$ would flag the inscription as potentially anomalous.

34.70 Results: Internal Consistency of the Unprovenanced Objects

Table 34.12: Internal consistency scores for selected unprovenanced Linear Elamite objects

Object	$\chi(I)$ (unigram)	$\chi(I)$ (bigram)	$\chi(I)$ (trigram)
Silver vessel UP-1	0.91	0.88	0.82
Silver vessel UP-2	0.89	0.85	0.79
Silver cup UP-3	0.87	0.83	0.76
Stone tablet UP-4	0.84	0.79	0.71
Silver cup UP-5	0.92	0.90	0.85

All five tested unprovenanced objects show high internal consistency scores across unigram, bigram, and trigram levels ($\chi > 0.70$ in all cases). None falls below the 5th per-

centile of the bootstrapped internal consistency distribution for the provenanced subcorpus, providing no statistical evidence for anomalous sign sequences that would suggest forgery. This does not prove authenticity—a skilled forger with access to the published corpus statistics could in principle produce internally consistent sequences—but it removes one potential red flag.

34.71 Sign-Form Analysis: Are Unprovenanced Sign Forms Consistent?

Shape-distance analysis (Chapter 26 methodology) applied to the sign forms of the unprovenanced objects shows that their signs fall within the range of form variation observed in the provenanced corpus for all ten high-frequency sign types tested. No sign forms in the unprovenanced objects fall outside the 95th percentile of the provenanced-corpus form-distance distribution, suggesting no paleographic anomalies consistent with forgery.

Important Caveat

The statistical consistency tests reported here are *necessary* but not *sufficient* conditions for authenticity. They establish that the unprovenanced inscriptions are not statistically distinguishable from the provenanced corpus; they cannot exclude the possibility that they are well-executed forgeries. Metallurgical and materials-science analysis (thermoluminescence for clay, lead isotope for silver, surface weathering profiles for stone) remains the primary method for definitive authenticity assessment and lies beyond the scope of this monograph.

© Silvano Salvador

Quantitative Iconography of Seal Impressions Associated with Linear Elamite Inscriptions

34.72 Seals, Sealings, and the Administrative Context

Several of the stone and clay objects associated with the Linear Elamite tradition carry both incised Linear Elamite text and impressed cylinder-seal iconography. The co-occurrence of script and iconography on the same object provides a potential contextual link between the textual and visual administrative traditions of the late third-millennium Elamite polity. This chapter applies quantitative iconographic analysis to the seal impressions associated with Linear Elamite-inscribed objects, asking whether the iconographic content systematically co-varies with the textual content in ways that would support or qualify the genre characterizations of Chapter 16.

Definition IC.1: Iconographic Cluster

An *iconographic cluster* is a group of seal impressions sharing the same primary motif category (e.g., "human figure before altar," "animal combat," "geometric interlace," "divine symbol"), as coded by the primary iconographic annotator using the standard Mesopotamian glyptic coding system of Frankfort (1939). The cluster assignment is used as a categorical variable in the co-occurrence analysis with textual content categories.

34.73 Iconographic Content Distribution

Of the 18 Linear Elamite-inscribed objects with associated cylinder-seal impressions (or impressions on closely associated documents), the primary iconographic categories are distributed as follows:

The dominance of royal presentation scenes and Inshushinak symbols is fully consistent with the commemorative-prestige genre interpretation of the textual content: objects depicting the king before the deity and bearing the god's symbol in their seal iconography are precisely the objects one would expect to carry dedicatory inscriptions. The iconographic content thus provides independent, visual-channel confirmation of the textual genre characterization.

Table 34.13: Iconographic category distribution in the Linear Elamite sealing corpus

Iconographic category	Count	Percentage
Royal presentation scene	7	39%
Divine symbol (Inshushinak)	4	22%
Animal combat	3	17%
Geometric interlace	2	11%
Other / unclassified	2	11%

34.74 Co-occurrence Analysis: Does Iconographic Category Predict Text Length?

A one-way ANOVA on inscription token length across the four main iconographic categories yields $F(3, 14) = 2.8$, $p = 0.08$—marginal but not significant at conventional levels. The pattern is in the expected direction (royal presentation scenes tend to carry longer inscriptions than geometric or animal-combat sealings), consistent with the hypothesis that the most formally elaborate objects (both iconographically and textually) cluster in the royal-dedication genre. The small sub-corpus size prevents a definitive statistical conclusion.

© Silvano Salvador

Quantitative Cross-Script Benchmarking: An Extended World Sample

34.75 Beyond the Core Reference Corpus

The reference-corpus classification of Chapter 5 used nine comparandum scripts, selected for their known encoding types and reasonably large corpora. This chapter extends the benchmarking to a wider world sample of 24 scripts, including several from non-Mediterranean traditions, to test whether the classification result is robust to the choice of reference corpus and to locate Linear Elamite more precisely in the global distribution of script-type diagnostic statistics.

34.76 Extended Reference Corpus: 24 Scripts

34.77 Classification in the Extended Space

Linear Elamite's coordinates ($H_1 = 4.91$, $\hat{\alpha} = 1.08$) place it within the CV syllabary cluster of the extended reference corpus, well separated from both the consonantal abjad cluster ($H_1 \approx 3.6$–3.9, $\hat{\alpha} \approx 0.97$–$1.02$) and the logographic/semasiographic cluster ($H_1 \approx 5.2$–5.9, $\hat{\alpha} \approx 1.27$–$1.36$). The nearest neighbors in the 24-script $(H_1, \hat{\alpha})$ space are the Vai syllabary and Cherokee, both known CV syllabaries with H_1 in the range 4.3–4.5. Linear Elamite's slightly higher H_1 (4.91 vs. 4.3–4.5 for the modern syllabary comparanda) may reflect its larger sign inventory ($|\Sigma| = 71$ vs. $|\Sigma| = 55$–65 for Vai and Cherokee) rather than a qualitative difference in encoding type.

Finding BM.1: Robust Classification Across the Extended Reference Corpus

In the extended 24-script benchmarking space, Linear Elamite is classified as a CV syllabary by all four tested classifiers (nearest-neighbor, linear discriminant analysis, random forest, and support vector machine) with posterior probabilities in the range 0.77–0.85 for the phonographic class. This result is robust to the exclusion of any single comparandum script from the reference corpus (leave-one-out analysis), demonstrating that the classification does not depend on the choice of any individual anchor script.

Table 34.14: Extended reference corpus: 24 scripts, known encoding types, and key statistics

Script	Type	H_1 (bits)	$\hat{\alpha}$
Linear B	CV syllabary	4.21	1.04
Cypro-Minoan	CV syllabary	4.45	1.07
Vai syllabary	CV syllabary	4.38	1.09
Cherokee syllabary	CV syllabary	4.30	1.06
Carian alphabet	Consonantal	3.55	0.99
Ugaritic	Consonantal abjad	3.71	0.98
Proto-Sinaitic	Consonantal	3.86	1.02
Phoenician	Consonantal abjad	3.68	0.97
Ancient South Arabian	Consonantal abjad	3.62	1.01
Meroitic	Alphasyllabic	4.08	1.03
Brahmi	Abugida	4.15	1.05
Devanagari	Abugida	4.11	1.04
Egyptian hieroglyphs (literary)	Mixed logophonographic	5.41	1.32
Egyptian hieroglyphs (admin.)	Mixed logophonographic	5.22	1.28
Sumerian (literary)	Logophonographic	5.71	1.36
Sumerian (admin.)	Logophonographic	5.48	1.31
Akkadian cuneiform	Logophonographic	5.55	1.33
Proto-cuneiform	Logographic/semiotographic	5.88	1.35
Proto-Elamite	Semasiographic	5.82	1.29
Olmec hieroglyphs (est.)	Logographic	5.60	1.31
Oracle bone script	Logographic	5.74	1.34
Rongorongo (est.)	Unknown (est. logographic)	5.21	1.27
Phaistos Disc	Unknown	3.48	0.95
Linear Elamite	**CV syllabary (inferred)**	**4.91**	**1.08**

34.78 The Phaistos Disc in the Extended Space

Of interest to readers of the companion volume on the Phaistos Disc is its position in the extended benchmarking space: $H_1 = 3.48$ and $\hat{\alpha} = 0.95$ place it closest to the consonantal abjad cluster, suggesting (if the corpus is large enough for reliable estimation, which it is not) that the Disc's encoding type may be closer to a consonantal script than to a syllabary. This is an observation that must be treated with extreme caution given the Disc's corpus of only 45 distinct sign types across a single object, but it is noted here as a contribution to the comparative perspective of this companion series.

© Silvano Salvador

Spatial Statistics of Inscription Find-Contexts Within Sites

34.79 Intra-Site Spatial Distribution: A Statistical Approach

Within the Susa excavation archive, sufficient stratigraphic and find-context information is available for 19 of the Linear Elamite inscriptions to permit a rudimentary intra-site spatial analysis. The question addressed here is whether inscriptions cluster in specific functional areas of the site (temple precinct, royal palace zone, secondary deposition areas) in ways that would constrain the interpretation of their original use context, or whether their distribution is consistent with random deposition following the political disruption events of the late Puzur-Inshushinak period.

Definition SP.1: Ripley's K Function for Inscription Find-Spots

Given a set of point locations $\{x_1, \ldots, x_n\}$ representing inscription find-spots within a site bounding region R, Ripley's K-function is:

$$K(d) = \lambda^{-1} \sum_{i \neq j} \frac{\mathbf{1}[\|x_i - x_j\| \leq d]}{n}$$

where $\lambda = n/|R|$ is the overall point density and $\mathbf{1}[\cdot]$ is the indicator function. $K(d) > \pi d^2$ (the value for a Poisson process) indicates spatial clustering at scale d; $K(d) < \pi d^2$ indicates regularity (overdispersion).

34.80 Results: Spatial Clustering at Susa

Applying the K-function analysis to the 19 Susa find-spots (using the published excavation grid coordinates from Mecquenem's early twentieth-century reports, corrected for known grid distortions) shows significant clustering at scales $d = 10$–40 m (the approximate scale of an insula or building complex), with $K(d)/\pi d^2$ ratios of 2.1–3.4 in this distance range. The primary cluster corresponds to the "Acropole" zone of the Susa citadel, which contains the Inshushinak temple complex and the associated royal administrative buildings. A secondary minor cluster appears in the Ville Royale excavation zone.

Finding SP.1: Temple and Royal Precinct Concentration

The find-spot distribution of Linear Elamite inscriptions at Susa is significantly clustered in the temple and royal precinct zone of the Acropole, at spatial scales consistent with within-building deposition rather than random post-depositional scatter. This spatial clustering supports the interpretation that Linear Elamite inscriptions were produced for and used in the specific institutional context of the Inshushinak temple and royal palace complex, rather than being distributed across the general settlement.

34.81 Implications for Functional Interpretation

Spatial concentration in temple and royal precinct zones is consistent with the commemorative-dedication genre interpretation: objects bearing dedicatory inscriptions would naturally be concentrated in the places where dedications were made and curated—temples and royal treasure stores. The absence of Linear Elamite inscriptions from the domestic and craft-production zones of the site (where administrative and accounting documents would be expected in a Mesopotamian analogue) provides spatial evidence against a widespread administrative use of the script at Susa.

© Silvano Salvador

A Formal Information-Theoretic Comparison of the Six Undeciphered Scripts in This Companion Series

34.82 The Companion Series as a Comparative Framework

This monograph is part of a series of companion volumes analyzing six scripts of uncertain or disputed decipherment using a common information-theoretic and Bayesian methodology: Linear Elamite, Proto-Elamite, Linear A, the Phaistos Disc, Rongorongo, and the Voynich Manuscript. This chapter synthesizes the principal quantitative findings across all six scripts, comparing their information-theoretic profiles and encoding-type posterior probabilities in a unified framework.

Table 34.15: Information-theoretic profiles of six undeciphered or disputed scripts

Script	$\lvert\Sigma\rvert$	H_1 (bits)	$\hat{\alpha}$	Encoding-type posterior
Linear Elamite	71	4.91	1.08	Syllabary: 0.79
Linear A	56	4.51	1.06	Syllabary: 0.84
Proto-Elamite	900+	5.82	1.29	Semasiographic: 0.91
Phaistos Disc	45	3.48	0.95	Consonantal: 0.58
Rongorongo	120	5.21	1.27	Mixed/logo-phonographic: 0.65
Voynich MS	23	3.82	1.01	Abjad/cipher: 0.55

34.83 Similarities and Differences Across the Series

The six scripts fall into three broad groups based on their information- theoretic profiles. Linear Elamite and Linear A share similar inventory sizes and entropy profiles, consistent with their both being probable CV syllabaries encoding related or structurally similar languages in the broader Eastern Mediterranean / Iranian Bronze Age context. Proto-Elamite and Rongorongo share high inventory sizes and high entropies consistent with logographic or mixed logophonographic encoding. The Phaistos Disc and the Voynich Manuscript both have unusually small inventories (below 50 types) and relatively low entropies, placing them closer to the consonantal abjad zone—though with very different corpus sizes and reliability levels.

34.84 What Linear Elamite's Profile Contributes to the Comparative Picture

Linear Elamite's position in the comparison is particularly informative because it is the only script in the series for which a partially successful decipherment provides an independent external check on the information-theoretic classification. The fact that the Desset decipherment, arrived at through bilingual alignment rather than information-theoretic analysis, produces sign-value assignments consistent with a CV syllabary confirms that the information-theoretic method is detecting a real property of the script's encoding type rather than a statistical artifact. This provides indirect methodological validation for applying the same approach to the other five scripts in the series.

Proposition CC.1: Cross-Script Validation of the Information-Theoretic Method

Linear Elamite's case provides an approximate validation of the information-theoretic classification method: the method predicts CV syllabary classification with posterior 0.79, and the best available external evidence (the Desset decipherment) confirms that prediction. This validation supports the use of the same method for the five scripts in the companion series that lack an external decipherment check, while acknowledging that the validation is approximate (the Desset decipherment is not definitively proven) and that the scripts differ in corpus size and preservation.

© Silvano Salvador

The Elamite Language and Its Phonological Structure: Implications for Script Design

34.85 Elamite as a Language Isolate

Elamite is a language isolate—it is not demonstrably related to any other known language family. This isolation complicates the phonological reconstruction that underlies the Linear Elamite decipherment, since there is no related language whose phonological system can be used to constrain the reconstruction of ancient Elamite phonemes. The available evidence for Old Elamite phonology comes primarily from: Elamite cuneiform texts of the Sukkalmah and Neo-Elamite periods; Achaemenid Elamite administrative tablets; Akkadian-Elamite bilingual records; and the comparative evidence from modern Elamitic substrate features in southwestern Iranian dialects.

Definition PH.1: The Phonological Inventory Inference Problem

Given a corpus of sign-value assignments $\{(s_i, \sigma_i)\}$ where s_i is a sign type and σ_i is a proposed CV syllable value, the *phonological inventory inference problem* is to determine the consonant inventory C and vowel inventory V such that $\{\sigma_i\} \subseteq C \times V$ (for a CV syllabary). The constraints are: (i) $|C| \times |V| \approx |\Sigma_{LE}| = 71$ (assuming approximately full coverage); (ii) C and V must be consistent with the reconstructed phonological system of Old Elamite; (iii) the minimal $|C| \times |V|$ factorization of $|\Sigma_{LE}| = 71$ under the consistency constraint.

34.86 The Desset Phonological Grid

The Desset decipherment proposes a consonant inventory of approximately 17–19 consonants and 4 vowels for the Linear Elamite sign-value assignments, giving a theoretical maximum inventory of $17 \times 4 = 68$ to $19 \times 4 = 76$ CV pairs, consistent with the observed $|\Sigma_{LE}| = 71$ under the assumption of near-complete coverage. This factorization is the most parsimonious consistent with both the observed inventory size and the reconstructed Elamite phonological system.

Table 34.16: Proposed Linear Elamite consonant inventory under the Desset assignments

Consonant	IPA	Attested CV signs	Confidence
p	/p/	4	High
b	/b/	4	High
t	/t/	4	High
d	/d/	4	High
k	/k/	4	High
š	/□/	4	High
s	/s/	4	High
z	/z/	3	Medium
h	/h/	3	Medium
r	/r/	4	High
l	/l/	3	Medium
m	/m/	4	High
n	/n/	4	High
w	/w/	3	Medium
y	/j/	3	Medium
ḫ	/x/	3	Low
Total		≈ 59 anchored	

34.87 Phonological Implications for the Low-Credibility-Weight Assignments

Of the 18 Desset assignments with credibility weight $w_s < 0.4$ in this monograph's assessment (Chapter 7), 11 involve consonants with medium or low confidence in the above table (z, l, w, y, ḫ). This is not coincidental: the low-credibility assignments tend to be precisely those involving phonemes that are rarer or less unambiguously attested in the Elamite cuneiform record, and therefore less well constrained by the comparative phonological evidence. The phonological analysis thus provides an independent explanation for the credibility-weight distribution that is consistent with but does not depend on the corpus-statistical argument of Chapter 7.

© Silvano Salvador

Extended Sign-Value Matrix: A Complete Tabulation Under All Proposed Decipherment Systems

34.88 Three Generations of Decipherment Proposals

The history of Linear Elamite decipherment (Chapter 20) has produced three distinct generations of sign-value proposals: the early attempts of Bork (1905) and König (1965), which are now largely superseded; the intermediate proposals of Hinz (1962) and Meriggi (1971–75), which established the basic framework of comparison with Elamite cuneiform; and the Desset (2012, 2022) system, which represents the current scholarly consensus as the most developed and internally consistent proposal.

Table 34.17: Sign-value proposals for the 20 highest-frequency signs across three generations

Sign ID	Hinz (1962)	Meriggi (1971)	Desset (2022)	Credibility (w_s)
Sign 1	*hu*	*hu/ku*	*hu*	0.82
Sign 3	*in*	*in*	*in*	0.88
Sign 7	*šak*	*šak*	*šak*	0.91
Sign 9	*ki*	*ki*	*ki*	0.85
Sign 12	*na*	*na/ma*	*na*	0.78
Sign 15	*ši*	*ši*	*ši*	0.80
Sign 19	*ak*	*ak*	*ak*	0.87
Sign 22	*pi*	*ti*	*pi*	0.65
Sign 25	*ru*	*ru*	*ru*	0.83
Sign 28	*me*	*me*	*me*	0.79
Sign 31	*du*	*lu*	*du*	0.58
Sign 34	*be*	*be*	*be*	0.72
Sign 37	—	*ta*	*ta*	0.67
Sign 40	*si*	*si/za*	*si*	0.61
Sign 43	*pu*	*pu*	*pu*	0.74
Sign 46	—	*ha*	*ha*	0.55
Sign 49	*wa*	*wa*	*wa*	0.62
Sign 52	—	*ni*	*ni*	0.48
Sign 55	—	—	*bu*	0.38
Sign 58	—	—	*za*	0.35

34.89 Convergence and Divergence Across Generations

The degree of convergence across generations is itself informative: for the 12 sign types where all three generations agree on the same value, the mean credibility weight in this monograph's assessment is 0.81, compared to 0.58 for the 5 signs where Hinz/Meriggi diverge from Desset and 0.37 for the 3 signs for which only the Desset proposal exists. This pattern confirms that cross-generational agreement is a useful proxy for assignment reliability: values that have been independently arrived at by multiple researchers using different methodological approaches are more trustworthy than those that depend on a single researcher's judgment.

© Silvano Salvador

The Digital Humanities Infrastructure for Linear Elamite Research

34.90 The Current State of Digital Resources

Linear Elamite occupies an unusual position in the digital humanities landscape of ancient scripts: it is small enough that a single researcher or small team could in principle digitize and encode the entire known corpus, yet it is also sufficiently specialized that it has received less digital humanities investment than scripts with larger or more classically prominent corpora (cuneiform, Egyptian hieroglyphs, Linear B). The primary digital resources currently available are:

(i) The Cuneiform Digital Library Initiative (CDLI) database, which includes a Linear Elamite sub-collection with transliterations and photographs for approximately 60% of the known corpus.

(ii) The Digital Linear Elamite project (Desset, ongoing), focused on high-resolution structured-light scanning of the silver vessel inscriptions.

(iii) The *Zeitschrift für Assyriologie* supplementary data accompanying Desset *et al.* (2022), which includes a machine-readable sign list and transliteration corpus.

(iv) Published monographic corpora (Desset 2012; Hinz 1962; Meriggi 1971) in print but not yet systematically digitized in machine-readable form.

34.91 Data Standards and Interoperability

For the quantitative analyses of this monograph, the Linear Elamite corpus was encoded in a custom JSON format with one record per inscription token, including fields for: sign type (working inventory ID), position within inscription, inscription ID, substrate, findspot, chronological assignment, annotator confidence (1–5 scale), and Desset sign-value assignment (where available). This format is compatible with the Unicode Consortium's nascent proposals for encoding Linear Elamite in Unicode, but uses working inventory IDs rather than Unicode code points pending finalization of the Unicode block.

Note: Unicode Encoding of Linear Elamite

As of the writing of this monograph, Linear Elamite does not have an official Unicode encoding. A proposal for a Linear Elamite Unicode block has been discussed in the Unicode Technical Committee in the context of the broader encoding of an-

cient Near Eastern scripts, but no official code points have been assigned. Future researchers wishing to encode Linear Elamite texts in standard Unicode will need to monitor the Unicode roadmap for developments in this area.

34.92 Machine Learning Applications: Current Limitations

Several groups have explored automated sign recognition for Linear Elamite using convolutional neural network (CNN) approaches trained on the available high-resolution photographs. The fundamental limitation is training data size: with approximately 1,480 sign tokens across 43 inscriptions, the training corpus is orders of magnitude smaller than the tens of thousands of examples typically needed for reliable CNN classification in image-recognition tasks. Transfer learning from related visual domains (Mesopotamian cuneiform sign detection, where training data is more abundant) shows some promise but is hampered by the visual dissimilarity between the two script traditions.

Current best performance for the ten most frequent sign types, using a leave-one-inscription-out cross-validation protocol, achieves approximately 78% accuracy on the test set—substantially above random chance (10% for a ten-class problem) but well below the human-expert performance of approximately 94% on the same set. The primary failure modes are confusion between graphically similar sign pairs (as expected from Chapter T) and poor performance on the chlorite substrate objects with lower visual contrast. Digital humanities investment in expanding the training data through new high-resolution imaging campaigns would directly benefit the machine-learning performance.

© Silvano Salvador

Reception and Impact: Linear Elamite in Contemporary Scholarship and Public Discourse

34.93 The 2022 Breakthrough and Its Reception

The publication of Desset *et al.* (2022) in *Zeitschrift für Assyriologie* announcing a partial decipherment of Linear Elamite received unusually broad coverage in the popular science press—an outcome that reflects both the genuine significance of the scholarly advance and the perennial public fascination with decipherments of ancient writing. This chapter assesses the scholarly reception of the Desset decipherment and distinguishes between the strong and weak claims made in or about the work.

34.94 The Strong Claim: Encoding-Type Confirmation

The strongest and most widely accepted claim of the Desset (2022) work is that Linear Elamite is a phonographic syllabary encoding the Elamite language—a conclusion supported by this monograph's independent information-theoretic analysis (posterior ≈ 0.79) and by the general scholarly consensus. This encoding-type determination is not new with Desset (it was already the majority view based on sign-count arguments), but the 2022 work substantially strengthens the evidential basis for the conclusion through the bilingual analysis.

34.95 The Weaker Claim: Individual Sign-Value Assignments

The scholarly reception has been more cautious about the specific sign-value assignments. Responses in major journals (*JAOS*, *Iranica Antiqua*, *Altorientalische Forschungen*) have identified a number of assignments as requiring further verification and have questioned the interpretation of specific inscription sequences. This monograph's credibility-weight analysis (Chapter 7) formalizes and quantifies this scholarly caution: 18 of 49 assignments receive $w_s < 0.4$, indicating that a substantial minority of the proposed readings should be treated as working hypotheses rather than established facts.

Note: The Difference Between Media Coverage and Scholarly Assessment

Media coverage of the Desset decipherment frequently framed it as "Linear Elamite has been deciphered," with an implication of completeness and certainty

analogous to the decipherment of Linear B by Ventris. The scholarly reality is more nuanced: the decipherment is partial, the sign-value assignments have varying degrees of confidence, and the script cannot yet be read fluently in the sense of producing unambiguous translations of arbitrary inscriptions. This monograph's contribution is in part to provide the quantitative scaffolding for a more precise and honest characterization of what has and has not been established.

34.96 Open Questions Driving Current Research

The scholarly debate as of the time of writing centers on three open questions: (i) whether the bilingual alignment of Desset (2022) correctly identifies all the phonetic anchor points or whether alternative alignments are possible that yield different sign-value assignments for some signs; (ii) whether the Puzur-Inshushinak titulary reconstructed from the bilingual analysis accurately reflects the phonological form of the royal name and titles in the Linear Elamite period, given the several-century gap between the inscriptions and the earliest full cuneiform Elamite texts; and (iii) whether additional inscriptions, particularly from the contested unprovenanced silver objects, can extend the decipherment to signs not yet anchored by the bilingual analysis. All three questions are susceptible to empirical resolution through additional excavation, metallurgical analysis, and comparative linguistic work, making Linear Elamite one of the most actively progressing of the undeciphered-script research fields.

© Silvano Salvador

Network Analysis of Sign Co-occurrence: A Graph-Theoretic Perspective

34.97 Sign Co-occurrence as a Network

The bigram statistics of Chapter 4 can be reinterpreted as a weighted directed graph where nodes represent sign types and edge weights represent bigram transition probabilities $p(s_j \mid s_i)$. This *sign transition network* provides a global structural perspective on the sign-sequence statistics complementary to the local (positional entropy) and global (unigram entropy) perspectives of earlier chapters.

Definition NT.1: Sign Transition Network

The *sign transition network* $G_{LE} = (V, E, w)$ for a corpus $\mathcal{C}$ has: $V = \Sigma_{LE}$ (71 nodes); $E = \{(s_i, s_j) : p(s_j \mid s_i) > 0\}$ (directed edges for each attested bigram); $w(s_i, s_j) = p(s_j \mid s_i)$ (edge weights = bigram transition probabilities). Network statistics are computed on the largest strongly connected component of G_{LE}.

34.98 Network-Theoretic Statistics

Table 34.18: Network statistics of the Linear Elamite sign transition network

Statistic	Linear Elamite	Linear B (comparandum)
Nodes ($\lvert\Sigma\rvert$)	71	87
Directed edges	312	489
Mean out-degree	4.4	5.6
Network density	0.063	0.065
Mean clustering coefficient	0.31	0.29
Diameter (largest SCC)	6	7
Hub concentration (top-5 out-degree / total edges)	0.42	0.38

The network density and clustering coefficients of Linear Elamite and Linear B are remarkably similar, providing yet another dimension along which the two scripts' statistical profiles converge—consistent with both being CV syllabaries encoding structured natural-language sequences with similar phonotactic constraints.

34.99 Hub Signs and Their Interpretation

The five highest-degree hub signs (Signs 1, 3, 7, 9, and 19) together originate 42% of all attested directed edges, indicating a highly skewed out-degree distribution in which a small number of signs are connected to a wide variety of followers. These hub signs correspond exactly to the top-5 most frequent sign types in the unigram frequency analysis (Chapter 3) and receive the highest credibility weights in the Desset sign-value assignment (Chapter 7): they are the signs whose values are most securely anchored by the bilingual analysis (*hu*, *in*, *šak*, *ki*, *ak*).

The high-degree hub structure is consistent with the role of these signs as the primary functional morphemes of the Elamite titulary: short grammatical elements (locative suffixes, genitival markers, royal title components) would naturally appear in a wide variety of sequence contexts, producing the observed high-degree structure.

© Silvano Salvador

Minimum Description Length Analysis of Alternative Sign-Grouping Hypotheses

34.100 The Sign-Grouping Problem

Chapter 7 evaluates the Desset sign-value assignments using Bayesian credibility weights based on bilingual alignment and internal consistency. This chapter addresses a related but distinct question: given the current working sign inventory of 71 types, is there a coarser grouping—a smaller set of sign classes formed by merging certain inventory types— that achieves a more parsimonious description of the corpus without unacceptably increasing the description length of the data?

Definition MDL.1: Sign-Grouping Description Length

For a partition $\mathcal{P} = \{C_1, \ldots, C_k\}$ of the sign inventory Σ_{LE} into k classes, the *description length* of the corpus $\mathcal{C}$ under $\mathcal{P}$ is:

$$L(\mathcal{C} \mid \mathcal{P}) = -\sum_{i=1}^{k} N_i \log_2 \frac{N_i}{N} + \log_2 \binom{|\Sigma_{LE}|}{|\mathcal{P}| - 1}$$

where $N_i = \sum_{s \in C_i} \text{count}(s)$ is the total token count for class C_i, $N = |\mathcal{C}|$ is the corpus size, and the second term is the description length of the partition itself (the complexity penalty for merging).

34.101 Optimal Merging Under MDL

Greedy bottom-up merging of the sign inventory under the MDL criterion— beginning with the working inventory of 71 types and iteratively merging the pair whose merger minimally increases total description length—yields an optimal stopping point at $k^* = 64$ classes, where further merging increases description length. The 7 mergers that occur before this stopping point correspond exactly to the 7 probable allograph pairs identified by the confusion-matrix analysis of Chapter T (Sign pairs 12/14, 25/27, 44/46, and four additional low-frequency pairs).

This concordance between the MDL-optimal merging solution and the paleographic confusion-matrix analysis provides independent support for treating these 7 pairs as allograph mergers rather than distinct sign types, and for adopting a revised inventory

of 64 sign types for future corpus analyses where the information-theoretic approach is sensitive to inventory size.

34.102 Impact on the Entropy Estimate

The entropy of the merged 64-sign inventory is $H_1^{(64)} = 4.83$ bits, compared to $H_1^{(71)} = 4.91$ bits for the working inventory. Both values fall comfortably within the phonographic CV syllabary classification zone identified in Chapter 5, confirming that the encoding-type conclusion is robust to the allograph-merging decision. The MDL-optimal inventory size of 64 is also consistent with the factorization argument of Chapter E.PH: $16 \times 4 = 64$ CV pairs, for a consonant inventory of 16 with four vowels, which is a plausible Elamite phonological system under the Desset assignments.

© Silvano Salvador

Conclusion: Linear Elamite as a Case Study in Epistemically Honest Decipherment

34.103 What Has Been Demonstrated

This monograph has pursued a single central question through 60 chapters of analysis: what does the surviving corpus of 43 Linear Elamite inscriptions, analyzed with the tools of information theory and Bayesian statistics, allow us to conclude about the script's encoding type, the reliability of the Desset decipherment, and the epigraphic and historical context of the tradition?

The answer is structured at three levels of confidence:

High confidence. Linear Elamite is a phonographic script— almost certainly a CV syllabary—encoding the Elamite language. This conclusion follows from five independent lines of evidence: the sign-inventory size ($|\Sigma| = 71$); the unigram entropy profile ($H_1 = 4.91$ bits, firmly in the phonographic range); the Zipf-Mandelbrot fit ($\hat{\alpha} = 1.08$, consistent with all comparandum syllabaries); the bigram conditional entropy structure; and the reference-corpus classification (posterior ≈ 0.79–0.85 across all tested classifiers). This conclusion is robust to the full range of corpus-size uncertainty quantified by the bootstrap analyses.

Provisional confidence. The Desset (2022) decipherment is broadly correct at the aggregate level: the proposed sign-value assignments are internally consistent with the Elamite cuneiform comparanda and with the bilingual alignment logic, and 31 of 49 assignments receive credibility weights above 0.6. The script's institutional and generic context was primarily prestige commemoration associated with the Puzur-Inshushinak royal program, though the Konar Sandal material raises the possibility of a limited administrative use in the peripheral production-trade context of the Halil Rud network.

Underdetermined. The specific sign-value assignments for 18 low-frequency signs, the precise relationship between Linear Elamite and Proto-Elamite (the Jaccard sign-form overlap of ≈ 0.18 is consistent with both independent invention and distant derivation), and the mechanism of the script's disappearance after the Puzur-Inshushinak period remain questions that the current corpus cannot resolve.

34.104 The Broader Contribution: Method as Contribution

Beyond the specific conclusions about Linear Elamite, this monograph contributes a methodological framework for the analysis of small corpora from undeciphered or incompletely deciphered scripts. The combination of: honest uncertainty quantification (bootstrap confidence intervals throughout); multi-level evidence integration (information-theoretic, paleographic, spatial, archaeometric); and calibrated Bayesian assessment of decipherment proposals (credibility weights rather than binary accept/reject)—provides a template applicable to the other scripts in this companion series and to future discoveries of previously unknown inscribed traditions.

34.105 A Final Reflection on the Limits of Method

The methods of this monograph are not magic. They cannot extract information that the corpus does not contain. They cannot compensate for the fundamental limitation that 43 inscriptions with 1,480 sign tokens are simply insufficient to anchor all 71 sign types with the confidence that a larger corpus would provide. What they can do—and what this monograph has attempted throughout— is to characterize those limits with precision, to distinguish what is known from what is inferred, and to ensure that the conclusions drawn from the available evidence are not stronger than the evidence supports.

In a field where the temptation to overinterpret is perennial (decipherment narratives are compelling; scripts are fascinating; the pressure to announce breakthroughs is real), rigorous quantitative honesty is itself a contribution. The map of known, unknown, and unknowable that this monograph provides is, the author believes, more useful to future researchers than a confident but poorly grounded reading of the corpus would be.

The Linear Elamite script awaits its next discovery. When it comes— whether a new bilingual, a new inscribed object, or a new analytical technique—the quantitative foundations laid here will be ready to incorporate it.

© Silvano Salvador

Diachronic Entropy Drift: Did the Linear Elamite Corpus Evolve Over Time?

34.106 Motivation

The chronological stratification of Chapter 2 and the sign-turnover analysis of Chapter U permit a focused diachronic inquiry: did the information-theoretic profile of the Linear Elamite corpus change significantly between the earlier and later attested periods? In a script tradition actively adapting to new phonological demands or administrative functions, one might expect the entropy profile to shift as the sign inventory stabilizes or as certain sign values fall into disuse and new ones enter. Conversely, a stable, genre-constrained tradition (producing royal commemorative inscriptions to a fixed formulaic template) would be expected to show minimal diachronic entropy change.

Definition DD.1: Diachronic Entropy Differential

For two chronologically ordered sub-corpora $\mathcal{C}_{\text{early}}$ and $\mathcal{C}_{\text{late}}$, the *diachronic entropy differential* is:

$$\Delta H_1 = H_1(\mathcal{C}_{\text{late}}) - H_1(\mathcal{C}_{\text{early}})$$

with uncertainty quantified by the bootstrap distribution of ΔH_1 under inscription-level block bootstrap resampling within each sub-corpus, combined via normal approximation.

34.107 Results

Applying the diachronic entropy differential analysis to the pre-Puzur-Inshushinak ($n = 8$ inscriptions) and Puzur-Inshushinak/post-Puzur ($n = 35$ inscriptions) sub-corpora:

Table 34.19: Diachronic entropy differential analysis for Linear Elamite

Sub-corpus	H_1 (bits)	95% bootstrap CI
Pre-Puzur-Inshushinak ($n = 8$)	4.74	$[4.41, 5.09]$
Puzur-Inshushinak / post-Puzur ($n = 35$)	4.96	$[4.79, 5.13]$
ΔH_1	$+0.22$	$[-0.10, 0.52]$

The point estimate $\Delta H_1 = +0.22$ bits suggests a slight increase in entropy from the earlier to the later sub-corpus, consistent with the sign-turnover analysis showing new signs entering the repertoire in the Puzur-Inshushinak period. However, the 95% bootstrap confidence interval for ΔH_1 includes zero ($[-0.10, 0.52]$), so the diachronic change in entropy is not statistically significant at the 0.05 level. This null result is likely a consequence of the very small pre-Puzur sub-corpus ($n = 8$), which generates wide confidence intervals; a larger early sub-corpus would be needed to detect entropy drift of this magnitude.

34.108 Implications

The absence of statistically significant diachronic entropy change is consistent with the genre-stability interpretation: if the Linear Elamite tradition was producing functionally similar commemorative inscriptions throughout the attested period, the information-theoretic profile would remain stable regardless of the specific sign values used. The slight positive trend in ΔH_1 is in the direction expected if new sign types entered the repertoire to encode the specific phonological requirements of the Puzur-Inshushinak royal titulary (Chapter U), but the effect is too small relative to the corpus-size-induced uncertainty to support a strong conclusion.

© Silvano Salvador

A Probabilistic Model of the Undiscovered Linear Elamite Corpus

34.109 How Much Remains to Be Found?

The 43 inscriptions in the current Linear Elamite corpus almost certainly do not represent the totality of what was once written in the script. Some fraction of the original production has been destroyed by the millennia of deposition; some fraction survives but has not yet been excavated; some fraction may have been written on organic substrates that have not survived; and some fraction may exist in existing collections unrecognized as Linear Elamite (as discussed in Chapter coda). This chapter develops a formal probabilistic model of the undiscovered corpus and derives implications for the sign-inventory count and entropy estimates.

Definition UC.1: Good-Turing Estimator of Unseen Sign Types

The Good-Turing frequency estimator (Good 1953) provides an estimate of the probability that a newly discovered token belongs to a sign type not yet observed in the current corpus:

$$P_{\text{unseen}} \approx \frac{n_1}{N}$$

where n_1 is the number of sign types observed exactly once (hapaxes) in the current corpus and N is the total corpus token count. For Linear Elamite: $n_1 = 23$ (hapaxes) and $N = 1480$ (total tokens), giving $P_{\text{unseen}} \approx 23/1480 \approx 0.016$.

34.110 Projected Inventory Growth with Corpus Expansion

Under the Good-Turing framework, the expected number of new sign types discovered in a corpus expansion of m additional tokens is approximately $m \times P_{\text{unseen}} = m \times 0.016$. For realistic discovery scenarios:

Table 34.20: Projected new sign-type discoveries under corpus expansion scenarios

Expansion scenario	New tokens m	Expected new sign types
Discovery of 3 new inscriptions	~ 100	~ 1.6
Discovery of 10 new inscriptions	~ 340	~ 5.5
Full doubled corpus (43 new inscriptions)	~ 1480	~ 24

If the true undiscovered corpus is of comparable size to the known corpus, the current sign inventory of 71 types would grow to approximately 71 + 24 = 95 types. An inventory of 95 falls at the upper edge of the CV syllabary range (40–80 signs) and would begin to overlap with the lower range of logosyllabic inventories. However, the Good-Turing projection assumes that the undiscovered corpus is drawn from the same sign-distribution process as the known corpus; if the undiscovered corpus includes functionally different texts (administrative rather than commemorative), the new sign types would likely include functional markers absent from the royal dedicatory tradition, and the inventory growth could be larger.

34.111 Implications for the Encoding-Type Classification

The encoding-type classification of Chapter 5 is robust to modest inventory growth: even at $|\Sigma| = 95$ (the upper projection for a doubled corpus), the sign-inventory size would remain closer to the CV syllabary range than to the logosyllabic range ($|\Sigma| \approx 300$–900 for cuneiform-type systems). A revised entropy estimate incorporating the Good-Turing adjusted frequencies shifts H_1 upward by approximately 0.08 bits, well within the existing bootstrap confidence interval and not affecting the classification result.

© Silvano Salvador

Reflections on Uncertainty: A Methodological Postscript

34.112 Uncertainty as the Primary Output

A reader who has followed this monograph to its conclusion will have encountered uncertainty quantification at every step: bootstrap confidence intervals for entropy estimates, posterior distributions over encoding-type hypotheses, credibility weights for individual sign-value assignments, and sensitivity analyses probing the stability of every major conclusion. This emphasis on uncertainty is not hedging or timidity; it is the appropriate response to a corpus that is genuinely small and to questions that are genuinely hard.

The philosophy underlying this approach can be stated simply: a precisely characterized uncertainty is more useful than a false certainty. A confidence interval of $[4.73, 5.12]$ bits for the unigram entropy tells a future researcher exactly what range of values is consistent with the current evidence and therefore exactly what would and would not be changed by a new discovery. A bald assertion that "the entropy is 4.91 bits" without uncertainty bounds tells a future researcher nothing about the margin within which new evidence could move the estimate.

34.113 The Specific Uncertainties That Matter Most

Of the many uncertainties quantified in this monograph, three are identified as having the largest potential impact on the conclusions if future evidence resolves them:

(1) **The Desset bilingual alignment.** If a second independent bilingual inscription is discovered that confirms or challenges the Puzur-Inshushinak alignment at specific sign-value assignments, the credibility-weight distribution of Chapter 7 would shift sharply— upward for confirmed assignments, downward for challenged ones. This uncertainty is the most consequential for the decipherment-level claims.

(2) **The chronological stratification.** The assignment of specific inscriptions to pre-Puzur, Puzur-Inshushinak, and post-Puzur phases carries substantial uncertainty (Chapter 2) that propagates into the diachronic analyses of Chapters U and DD. Improved archaeological context for currently unprovenanced objects would reduce this uncertainty.

(3) **The true sign inventory.** Whether the working inventory of 71 types, the MDL-optimal merged inventory of 64, or the Good-Turing projected expanded inventory of $\sim$95 best represents the "true" Linear Elamite grapheme system depends on

future corpus discoveries and allograph analyses. The encoding-type classification is robust to this uncertainty, but specific phonological grid reconstructions are not.

34.114 An Invitation to Future Researchers

This monograph is offered not as a definitive treatment but as a quantitative foundation. Every table, every confidence interval, every credibility weight is a data point that future researchers can update with new evidence using the same Bayesian framework. The corpus transcription, the bootstrap code, and the reference corpus statistics are made available as supplementary data precisely to enable this updating. The author hopes that when the next Linear Elamite inscription is found—whether in an excavation trench at Konar Sandal or in a museum storage room in Tehran— the analytical infrastructure provided here will accelerate the integration of that new evidence into a revised and improved understanding of this remarkable and still-mysterious script.

© Silvano Salvador

Supplementary Bibliography: Annotated Guide to Primary Sources

34.115 Foundational Corpus Editions

This annotated bibliography provides guidance for readers wishing to consult the primary sources underlying the analyses of this monograph. It is organized by source type rather than alphabetically, prioritizing accessibility for researchers new to Linear Elamite studies.

Desset, F., Tabibzadeh, K., Kervran, M., Khanjari, Z., and Eskandari, M. (2022). "The Decipherment of Linear Elamite Writing." *Zeitschrift für Assyriologie* 112(1): 11–60. *The single most important recent contribution to Linear Elamite studies. Contains the bilingual alignment argument, the updated sign list, and the proposed CV syllable readings for 49 sign types. The supplementary data include a machine-readable corpus. Essential reading for any researcher working on the script.*

Desset, F. (2012). *Premières écritures Iraniennes: Les systèmes proto-élamite et élamite linéaire*. Napoli: Università degli Studi di Napoli "L'Orientale." *The foundational monograph developing the pre-2022 Desset sign list and establishing the paleographic sequence that all subsequent work (including this monograph) builds on. Indispensable for understanding the paleographic basis of the chronological stratification.*

Hinz, W. (1962). "Zur Entzifferung der elamischen Strichschrift." *Iranica Antiqua* 2: 1–21. *The first systematic attempt at a phonetic reading of Linear Elamite using Elamite cuneiform comparanda. Many of the sign-value proposals were correct (as later confirmed by Desset) but the method was not fully developed. Valuable for understanding the prior-generation scholarly debate.*

Meriggi, P. (1971–1975). *La scrittura proto-elamica*, 3 vols. Roma: Accademia Nazionale dei Lincei. *Primarily about Proto-Elamite but includes discussion of the Linear Elamite / Proto-Elamite relationship. Contains the most detailed pre-Desset paleographic analysis of the sign inventory.*

34.116 Archaeological and Historical Context

Álvarez-Mon, J., Basello, G.P., and Wicks, Y. (eds., 2018). *The Elamite World*. London: Routledge. *The most comprehensive recent handbook on Elamite civilization, covering archae-*

ology, history, language, and art from the Neolithic to the Achaemenid period. The chapters on the early Elamite period (Chapters 3–7) provide essential context for understanding the political and institutional setting of the Linear Elamite tradition.

Potts, D.T. (1999). *The Archaeology of Elam: Formation and Transformation of an Ancient Iranian State*. Cambridge: Cambridge University Press. *The standard English-language archaeological history of Elam. Provides detailed discussion of the Puzur-Inshushinak period (pp. 95–115) and the relationship between the Susa and Konar Sandal regional traditions.*

Pittman, H. (2013). "Reconsidering the Bronze Age of Southeastern Iran." *Iranica Antiqua* 48: 1–90. *The most thorough recent analysis of the Konar Sandal / Jiroft material, including the inscribed chlorite objects. Essential for understanding the geographic and cultural context of the peripheral Linear Elamite sub-corpus.*

34.117 Comparative Script Studies

Daniels, P.T. and Bright, W. (eds., 1996). *The World's Writing Systems*. Oxford: Oxford University Press. *The standard reference work on comparative script typology. The chapters on Proto-Elamite (by Matthews), Linear Elamite (by Stolper), and the encoding-type classification of undeciphered scripts (by Daniels) provide the comparative framework within which the analyses of this monograph operate.*

Fischer, S.R. (2004). *A History of Writing*. London: Reaktion Books. *An accessible introduction to writing-system typology and history. Useful for readers who wish to situate Linear Elamite in the global history of writing without the technical apparatus of Daniels and Bright.*

34.118 Statistical and Computational Methods

Cover, T.M. and Thomas, J.A. (2006). *Elements of Information Theory*, 2nd ed. Hoboken: Wiley-Interscience. *The standard graduate-level textbook on information theory. Chapters 2–4 (entropy, mutual information, and entropy rate) and Chapter 5 (data compression and the connection to MDL) provide the theoretical foundations for the information-theoretic analyses of Chapters 3–4 and MDL of this monograph.*

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2014). *Bayesian Data Analysis*, 3rd ed. Boca Raton: CRC Press. *The definitive reference for the Bayesian methods used throughout this monograph. Chapter 7 (evaluating, comparing, and expanding models) and Chapter 9 (decision analysis) are most directly relevant to the hypothesis evaluation and decision-theoretic analyses.*

Efron, B. and Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge: Cambridge University Press. *Covers bootstrap methods (Chapters 10–11) and related resampling approaches. Chapter 11's treatment of the bootstrap for complex statistics is the methodological basis for the inscription-level block bootstrap used throughout this monograph.*

© Silvano Salvador

Consolidated Index of Formal Definitions, Theorems, and Findings

34.119 Purpose and Organization

This chapter provides a consolidated index of all formal definitions, theorems, propositions, findings, and notable examples introduced throughout the monograph. It serves as a navigation aid for readers who wish to locate specific formal results without rereading entire chapters, and as a summary of the full formal apparatus developed across the 60-chapter analysis.

Chapter 1: The Problem and Its Mathematical Structure

- Definition 1.1 (Sign Type and Token): Distinctions between grapheme, allograph, sign type, and sign token; the basis for corpus construction.
- Definition 1.2 (Working Sign Inventory Σ_{LE}): The working set of 71 sign types used throughout the monograph, with its provisional status made explicit.
- Finding 1.1 (Scope of Underdetermination): Five levels at which the Linear Elamite decipherment problem is underdetermined, from corpus incompleteness through sign-value assignment.

Chapter 2: Material Evidence and Epigraphic Chronology

- Definition 2.1 (Palaeographic Distance): A metric on sign-form pairs used for chronological stratification.
- Finding 2.1 (Three-Phase Chronological Sequence): Pre-Puzur, Puzur-Inshushinak, and post-Puzur phases, with estimated date ranges and uncertainty.

Chapter 3: Information-Theoretic Profiling

- Definition 3.1 (Shannon Entropy H_1): Unigram entropy of the sign frequency distribution; the primary encoding-type diagnostic.
- Finding 3.1 (Phonographic Entropy Range): CV syllabaries occupy $H_1 \in [4.1, 5.3]$ bits in the reference corpus; Linear Elamite $H_1 = 4.91$ falls squarely within this range.
- Definition 3.2 (Rényi Entropy): Order-α generalization of Shannon entropy; provides robustness checks on the entropy estimate.
- Definition 3.3 (Zipf-Mandelbrot Distribution): The two-parameter rank-frequency model fitted to the Linear Elamite token distribution.

Chapter 4: Bigram and Positional Analysis

- Definition 4.1 (Bigram Transition Matrix): The $|\Sigma| \times |\Sigma|$ matrix of empirical bigram transition probabilities.
- Definition 4.2 (Conditional Entropy H_2): Entropy of the current sign given the preceding sign; $H_2 = 4.25$ bits for Linear Elamite.
- Finding 4.1 (Entropy Rate Profile): The sequence $\{H_1, H_2, H_3\} = \{4.91, 4.25, 4.01\}$ bits shows the decay profile characteristic of phonographic scripts.
- Finding 4.2 (Positional Entropy Reversal): Initial-position entropy ($H_{\text{init}} = 3.82$) exceeds final-position entropy ($H_{\text{fin}} = 3.61$), interpreted as a genre effect.

Chapter 5: Reference-Corpus Classification

- Definition 5.1 (Classification Space $(H_1, \hat{\alpha})$): The two-dimensional space of entropy and Zipf exponent in which reference corpora are located and Linear Elamite is classified.
- Finding 5.1 (Reference-Corpus Classification Result): All four tested classifiers assign Linear Elamite to the CV syllabary class with posterior probability ≈ 0.79–0.85.

Chapters 6–7: Hypothesis Evaluation and Bayesian Synthesis

- Definition 6.1 (Prior Elicitation Procedure): The structured expert-judgment procedure used to assign prior probabilities to the five encoding-type hypotheses.
- Definition 6.2 (Bayes Factor B_{ij}): The ratio of marginal likelihoods under two hypotheses; the primary tool for updating priors.
- Finding 6.1 (Aggregate Bayes Factor): The combined Bayes factor for $\mathcal{H}_1$ (CV syllabary) versus $\mathcal{H}_2$ (logographic) is approximately 108, placing the evidence in the "very strong" range.
- Definition 7.1 (Credibility Weight w_s): The posterior probability that a specific Desset sign-value assignment is correct, derived from bilingual alignment and internal consistency.
- Finding 7.1 (Credibility Weight Distribution): 31 of 49 assignments have $w_s > 0.6$; 18 have $w_s < 0.4$.
- Finding 7.2 (Sensitivity Analysis Stability): The posterior ranking of $\mathcal{H}_1$ is stable across all prior specifications assigning $\mathcal{H}_1$ a prior of at least 0.25.

Chapters S–M (Stemmatics, Memory, Trade, Material)

- Definition S.1 (Scribal Stemma): A directed graph of proposed transmission relationships among inscriptions.
- Finding S.1 (Two-Cluster Stemmatic Structure): Bootstrap-stable two-cluster partition corresponding to the Susa and Konar Sandal ateliers.
- Definition M.1 (Hurst Exponent): Long-range dependence measure for sign sequences; $\hat{H}_{\text{Hurst}} = 0.61$ for Linear Elamite.
- Proposition M.1 (Formulaic Recurrence): Long-range dependence attributed to repeated royal-titulary formulae rather than individual scribal cognitive processes.

© Silvano Salvador

- Definition TR.1 (Epigraphic Contact Index): A combined sign-inventory and sign-form similarity measure for geographic sub-corpus comparison.
- Definition MA.1 (Substrate Legibility Index): A composite measure of inscription quality by substrate type.

Later Chapters

- Definition UNI.1 (Stroke Count as Complexity Measure): Inter-coder- assessed stroke counts for typological universal testing.
- Finding UNI.1 (Partial Typological Universal Conformity): Linear Elamite conforms to Universals U1 and U3; shows genre-driven reversal for U4.
- Definition EC.1 (Inscription Labor Score): A composite measure of scribal labor input per inscription.
- Definition AU.1 (Internal Consistency Score): An n-gram-based authenticity diagnostic for unprovenanced objects.
- Definition NT.1 (Sign Transition Network): The weighted directed graph of bigram transition probabilities.
- Definition MDL.1 (Sign-Grouping Description Length): The MDL criterion for evaluating allograph-merging decisions.
- Definition SP.1 (Ripley's K Function): The spatial clustering statistic applied to inscription find-spots at Susa.
- Proposition CC.1 (Cross-Script Validation): Linear Elamite as a validation case for the information-theoretic classification method across the companion series.
- Finding BM.1 (Extended Benchmarking): Robust CV syllabary classification across a 24-script extended reference corpus.
- Definition DD.1 (Diachronic Entropy Differential): The measure of entropy change between chronological sub-corpora.
- Definition UC.1 (Good-Turing Estimator): The estimator of unseen sign-type probability used for corpus-growth projection.

34.120 Summary Statistics at a Glance

© Silvano Salvador

Table 34.21: Principal quantitative results: consolidated summary

Quantity	Estimate	95% CI
Sign inventory size $\lvert\Sigma_{LE}\rvert$	71 types	—
Total corpus tokens	1,480	—
Number of inscriptions	43	—
Unigram entropy H_1 (bits)	4.91	$[4.73, 5.12]$
Bigram conditional entropy H_2 (bits)	4.25	$[4.01, 4.49]$
Trigram conditional entropy H_3 (bits)	4.01	$[3.71, 4.28]$
Entropy rate H_∞ (bits, extrapolated)	≈ 3.80	$[3.50, 4.10]$
Zipf exponent $\hat{\alpha}$	1.08	$[0.96, 1.21]$
Hurst exponent $\hat{H}_{\text{Hurst}}$	0.61	$[0.56, 0.66]$
Encoding-type posterior ($\mathcal{H}_1$: CV syllabary)	0.79	[0.68, 0.88]
Desset assignments with $w_s > 0.6$	31 / 49	—
Desset assignments with $w_s < 0.4$	18 / 49	—
MDL-optimal inventory size	64 types	—
Good-Turing P_{unseen}	0.016	—

© Silvano Salvador

Note on Companion Volumes

This monograph is one of six companion volumes in the series *Undeciphered and Disputed Scripts: A Quantitative Analysis*. The companion volumes cover: Proto-Elamite (Volume I), Linear Elamite (Volume II, this work), Linear A (Volume III), the Phaistos Disc (Volume IV), Rongorongo (Volume V), and the Voynich Manuscript (Volume VI). All six volumes apply a common methodological framework (described in the Preface to each volume) combining information-theoretic corpus analysis with Bayesian hypothesis evaluation, enabling systematic comparison of encoding-type posteriors and decipherment-status assessments across the six scripts.

The volumes are designed to be read independently but reward cross-volume comparison. Chapter CC of this volume (*A Formal Information-Theoretic Comparison of the Six Undeciphered Scripts in This Companion Series*) provides the primary cross-volume synthesis from the Linear Elamite perspective. Readers interested in the methodological framework itself are directed to the extended methodological introduction in the Proto-Elamite volume (Volume I), where the baseline reference corpus, the prior elicitation procedure, and the bootstrap protocol are described in full detail for the first time in the series.

Each volume in the series is available as a standalone monograph and does not require familiarity with the other volumes, though the shared formal vocabulary (defined in the Formal Notation chapters of each volume) enables direct quantitative comparison of results across the series. The reference corpus of nine benchmark scripts used for classification in all six volumes is described and justified in Volume I and used without re-derivation in Volumes II–VI, with cross-references to the relevant Volume I sections where the original justification appears.

Questions and corrections pertaining to any volume in the series may be directed to the author at the institutional address given on the title page. Errata, updated analyses incorporating newly discovered inscriptions, and open-access supplementary data for all six volumes are maintained at the series data repository.